Realized Copulae in Moderate Dimensions

Master Thesis submitted to

Prof. Dr. Ostap Okhrin

Prof. Dr. Wolfgang K. Härdle

Institute for Statistics and Econometrics
Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. - Center for Applied Statistics and Econometrics

Humboldt-Universität zu Berlin

by

Tatjana Tissen-Diabaté

(541353)

in partial fulfillment of the requirements

for the degree of

Master of Sciences in Statistics

Berlin, December 9, 2013

Abstract

For modeling and quantifying the dependence structure between financial risk factors like stock returns, copula models are a useful instrument.

Due to technological growth and improvements of the last decades, high-frequency data with intra-daily observations can be used to estimate covariance of assets. From these estimates and assumed marginal distribution, realized copula models are constructed and analyzed using real price data.

Based on high-frequency data, three different estimation methods are applied and in addition daily stock return are used to get Maximum Likelihood estimates. All estimators are implemented and the resulting copula models are compared according to their ability to describe the true multivariate distribution and forecast the losses.

The modeled dependence parameter are time-varying and differ according to the estimated covariance matrix and to the sampling frequency. Moreover, all estimation steps are done based on data with high sampling frequencies like seconds and lower frequency of 5 minutes. The realized kernel estimator is used to cope with market microstructure effects and the sensitivity to this noise can be reflected in the Value-at-Risk performance.

Keywords:

High frequency data, realized copula, realized kernel, realized volatility HAR model, Value-at-Risk, market microstructure noise

Zusammenfassung

Copulae sind ein hilfreiches Werkzeug um die Abhängigkeitsstrukturen zwischen Risikofaktoren in Wirtschaft und Finanzwesen zu modellieren und zu quantifizieren. Dank technologischen Entwicklungen der letzten Jahrzehnte, können hoch-frequentierte Daten mit innertäglichen Beobachtungen genutzt werden um die Covarianz zwischen Wertpapieren zu schätzen. Aus dieser Schätzung und zusätzlichen Annahmen über die Randverteilungen werden in dieser Arbeit Realized Copula Modelle an Hand von Aktienkursen auf verschiedenen Wegen geschätzt und untersucht.

Basierend auf hoch-frequentierten Daten werden drei Schätzer für den Copula Parameter angewendet und durch die Nutzung von täglichen Renditen wird ein Maximum Likelihood Schätzer bestimmt. Die unterschiedlich geschätzten Copula parameter werden verglichen und die resultierenden Copula Modelle bezüglich ihrer Angepasstheit an die Daten und ihrer Vorhersage von Verlusten untersucht.

Die modellierte Abhängigkeit ändert sich über die Zeit und unterscheidet sich je nach verfügbarer Kovarianzschätzung und Ziehungshäufigkeit der innertages Stichproben. Dabei werden die Realized Kernel Schätzer, bei sehr häufiger Ziehung, zum Teil bei der Schätzung der Copula Parameter genutzt. Einige Parameterschätzer liefern präzisere Quantilsschätzungen der Verlustverteilung als andere, was durch eine Value-at-Risk Schätzung und durch backtesting gezeigt wird.

Schlagwörter:

Hochfrequentierte Daten, Realized Copula, Realized Kernel, Realized Volatility, HAR Modell, Value-at-Risk, Market Microstructure Noise

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Understanding and modeling dependence of stock prices is a central concern in financial econometrics. Occasionally, abrupt and unexpected changes in the price dynamics occur in times of financial crisis, e.g. the ongoing European sovereign dept crisis during the last years. Therefore, in the last few decades much research was done on evaluating risk factors and dependence structures between different assets.

New models that better fit the dynamics of risk factors are developed that are useful for risk management, asset pricing and other areas. Unfortunately, these are often simulation-based methods or numerically intensive and they make use of non-parametric statistics. Rank statistics, for example, provide a simple tool for estimation when the linear correlation is not appropriate. This is mostly the case with real financial data, since the assumption of a normal distribution is often rejected.

Several years ago, the multivariate Gaussian distribution was frequently used for modeling and analyzing risk across many assets until in times of crisis assets revealed higher dependence especially higher in times of large crashes (Oh and Patton (2011)) than in times of boom. The multivariate Gaussianity does not allow for tail dependence and therefore does not capture joint movement in extreme events. Departure from the Gaussian distribution implies that the linear correlation coefficient is not sufficient for capturing all the dependence between assets.

One new approach, which finds much attraction recently and that is able to measure the complete dependence structure, are copula models. This is due to their distinct property of separating estimation of the multivariate distribution from the estimation of the marginal distributions. In practice, there is often much more information about the marginal distributions than about the multivariate structure. Copulae facilitate the estimation of the joint distribution by capturing the multivariate structure inside the copula dependence parameter. There are different estimation methods, like a maximum likelihood estimation. However, a new trend in copula estimation has appeared in the last years, which uses more information from intraday observations. Great developments in computer technology at the beginning of the 21st century enabled researchers to work with large data amounts. Standard estimation methods cannot be applied in this field. So computations based on high frequency intraday data developed new realized measures and there is much empirical literature concerned with realized volatility modeling for example Barndorff-Nielsen and Shephard (2004) and Corsi (2009). The realized variance is an estimator for volatility based on high-frequency data that is easy to compute. However, in presence of market microstructure noise this estimator is not reliable. The higher the frequency of intraday sampling the more noise can be assumed. In this work the realized kernel estimator will be used alternatively to the simple realized volatility estimators.

Since the covariation estimates are based on high-frequency data, the constructed copula model

based on these estimates is called realized copula. This concept was introduced by Fengler and Okhrin (2012) and it is quite new.

Beside measuring the variation of individual assets, the dependence structure of financial risk factors is estimated via copulae. Basically, copulae are multivariate distributions that model simultaneously a number of random variables.

There are several important qualities of copulae for practical applications

1. When the marginal distributions are constructed in the first step, they can belong to different parametric families or even be non-parametric. This is very useful in practice for financial market data where the marginals exhibit high kurtosis and skewness.

2. Copulae summarize the dependence structure of random variables in a dependence parameter, that is often one-dimensional and permits the researcher to examine and analyze the observed dependence structure.

3. Attributes of non-normality, like tail-dependence, skewness and asymmetric dependence, can be modeled so that properties of stock returns often observed in practice are reflected in the model.

Though it is not that easy to find a copula model that will work well in practice. In the last years a number of new and flexible models was developed (see McNeil et al. (2005), Oh and Patton (2011))

In general, there are three important dependence concepts: linear correlation, rank correlation and tail dependence. Multivariate structure and dependence measure are detailed in the next chapter. In the third chapter the notion of copula models is introduced together with some examples of copulae, estimation methods and the concept of realized copula.

In the subsequent forth chapter, data from three stocks is analyzed relating to the covariation and dependence structures. It is shown, that estimates of the copula parameter based on the intraday data at hand delivers better estimation results, compared to estimation with daily data. Also estimates for the Value-at-Risk are given and forecasts are made based on different estimation methods for the copula dependence parameter.

Chapter 2

Multivariate Distributions and Dependence Concepts

Copulae, as a special form of multivariate distribution, play a great role in the study of dependence. In Joe (1997) the terms dependence structure and multivariate structure are used equivalent. Therefore, some theory on multivariate distributions and subsequently on dependence concepts is presented in this chapter.

Let $X = (X_1, \ldots, X_d)$ be a random vector, beside the univariate distributions further information about the multivariate structure of X is needed for modeling X. For the marginal distributions $F_i(x_i) = P(X_i \leq x_i)$ for $i \in \{1, 2, \ldots, d\}$, described in the following, estimates can be obtained either by parametric or by nonparametric methods. In the parametric case, maximum likelihood estimation can be performed to get estimates for the parameters. Nonparametrically, a marginal distribution function can be approximated with the empirical cumulative distribution function

$$\hat{F}_n(x) = \frac{1}{T+1} \sum_{i=1}^{T} \mathbf{I}\{x_i \leq x\}$$

where T is the number of observations of a sample and $\mathbf{I}$ is the indicator function with

$$\mathbf{I}\{x_i \leq x\} = \begin{cases} 1 & \text{when } x_i \leq x \\ 0 & \text{when } x_i > x \end{cases}$$

For a univariate cumulative distribution functions F it holds that F is monotone increasing, that is $F(x_2) - F(x_1) >= 0$ for all $x_1, x_2 \in R$ such that $x_1 <= x_2$, and bounded between zero and one with $F(-\infty) = 0$ and $F(\infty) = 1$. This can easily be transferred to the bivariate case. A bivariate distribution function F satisfies

1. $\lim_{x_i \to -\infty} F(x_1, x_2) = 0$ for $i = 1, 2$ and $\lim_{x_i \to \infty \forall i} F(x_1, x_2) = 1$
2. Rectangle inequality: $F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2) \geq 0$ for all $(a_1, a_2), (b_1, b_2)$ with $a_1 < b_1$, $a_2 < b_2$.

2.1 Multivariate Distribution Function

In the past, the multivariate normal distribution was frequently used for empirical applications. However, it is not a good characterization of financial risk factor returns (McNeil et al. (2005)). Non-normal distributions form an area of probability and statistics of increasing interest. In the beginning of the 21st century, research in the area of multivariate non-normal distributions was intensified (Joe (1997)). Properties of a general multivariate cumulative distribution are presented in the following.

A d-dimensional cumulative distribution function of a random vector $X = (X_1, \ldots, X_d)$ is a right-continuous function F on $\mathbb{R}^d$, given as

$$F(\mathbf{x}) = F(x_1, \ldots, x_d) = P(X_1 \leq x_1, \ldots, X_d \leq x_d)$$

which satisfies the following conditions.

1. The distribution function F is increasing in each component x_i with the extreme cases: $\lim_{x_i \to -\infty} F(x_1, \ldots, x_d) = 0$ and $\lim_{x_i \to \infty \forall i} F(x_1, \ldots, x_d) = 1$ for $i = 1, \ldots, d$.

2. For all $(a_1, \ldots, a_d)$, $(b_1, \ldots, b_d) \in [0,1]^d$ with $a_i \leq b_i$, $i = 1, \ldots, d$ the rectangle inequality

$$\sum_{i_1=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{i_1 + \cdots i_d} F(x_{1,i_1}, \ldots, x_{d,i_d}) \geq 0$$

   is fulfilled, where $x_{j1} = a_j$, $x_{j2} = b_j$ for all $j = 1, \ldots, d$.

Related to the density function f the multivariate distribution can be written as

$$F_X(x_1, \ldots, x_d) = \int_{-\infty}^{x_d} \cdots \int_{-\infty}^{x_1} f_X(u_1, \ldots, u_d) du_1 \ldots du_d$$

The joint density function $f_X(\mathbf{x})$ with $\mathbf{x} = (x_1, \ldots, x_d)^\top$ returns positive values $f_X(\mathbf{x}) > 0$ for all $x_i \in \mathbb{R}$ and "sums up to 1":

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_X(x_1, \ldots, x_d) dx_1 \ldots dx_d = 1$$

From the joint cumulative distribution function the marginal distribution can be obtained by integrating out all other random variables. For the bivariate case, the marginals density function F_1 can be obtained from $F(x_1, x_2)$ when $F(x_1, \infty) = P(X \leq x_1, X_2 \leq \infty) = P(X_1 \leq x_1) = F_1(x_1)$.

Considering a multivariate distribution, for multivariate modeling the dependence structure plays a major role. The following sections present three different kinds of dependence measures which can be applied for a pair of random variables (X_1, X_2) and returns a skalar measure.

2.2 Dependence Measures

For modeling multivariate data, assumptions on the dependence structure between the random variables need to be made. The concept of dependence is examined by measures of association and it is one of the most widely studied subjects in statistic. Many of these measures are invariant to strictly

increasing transformations (Nelsen (2007)). Dependence concepts can vary in type and range of dependence. Often some range of the parameters correspond to positive or negative dependence. For some parametric families some of the parameters can be identified as dependence (or multivariate) parameters, for example the covariance matrix for a multivariate normal distribution (Joe (1997)). The independence is a special case, where the joint distribution can be expressed as a product of the marginal distributions.

For a random vector $X = (X_1, \ldots, X_d)^\top \in \mathbb{R}^d$ the joint distribution F can be determined by the probabilities

$$P(a_1 \leq X_1 \leq b_1, \ldots, a_d \leq X_d \leq b_d), \quad -\infty < a_i \leq b_i < \infty,\ i = 1, \ldots, d$$

When all components of X are mutually independent, the probability can be written as

$$P(a_1 \leq X_1 \leq b_1, \ldots, a_d \leq X_d \leq b_d) = \prod_{i=1}^{d} P(a_i \leq X_i \leq b_i)$$

For dependent random variables, it is more difficult to quantify the dependence structure. Kendall's tau and Spearman's rho are the well-known and most commonly used measures of association, beside the standard Pearson's linear correlation coefficient. However, for non-normal random variables the linear correlation is not a good dependence measure, as addressed in numerous literature, see e.g. Joe (1997). Reasons for this and basic properties are presented in the following.

2.2.1 Linear Correlation

Normal or, more generally, elliptical distributions are fully described by a mean vector, a covariance matrix and a characteristic generator function. That is why Pearson's linear correlation coefficient is useful for elliptical distributions for which it can capture the dependence between two different variables by using information from the covariance matrix.

The linearity of this dependence measure gets clear by stating the following linear relationship

$$Y = a + bX, \quad with\ a, b \in \mathbb{R}$$

where a correlation of 1, for instance, is achieved if $a = 0$ and $b = 1$.

If two random variables X and Y have a joint normal distribution, then the covariance contains the information about the dependence between these variables. The linear correlation coefficient is defined as

$$\rho = Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} \tag{2.1}$$

where $Cov(X, Y) = E\{(X - \mu_X)(Y - \mu_Y)\} = E(XY) - E(X)E(Y)$ is the covariance of X and Y. As usual, $\mu_Z = E(Z)$ denotes the mean and $Var(Z) = E(Z^2) - \{E(Z)\}^2$ the variance of a random variable Z .

The linear correlation has the property of taking values between -1 and 1. These bounds represent perfect negative and positive correlation, respectively. In figure 2.2.1 simulated values from a normal distribution are shown with a linear correlation coefficient of 0.7 for the positive and -0.7 for the negative dependence.

Only for the normal joint distribution a correlation of zero is equivalent to independence. Regarding the other way, for independent pairs of random variables of any other distribution, it holds that the

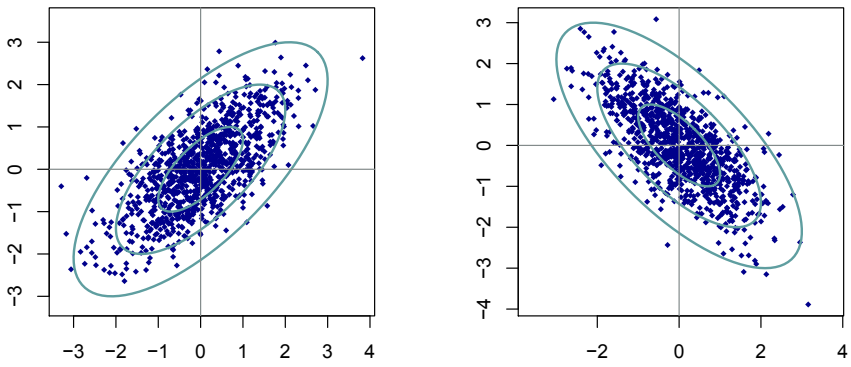

Figure 2.2.1: Two standard normal distributed variables with correlation 0.7 (left) and -0.7 (right). Kendall's tau takes the values -0.49 and 0.48 for the left and right case, respectively. Spearman's rho is very near to Pearson's correlation, namely -0.68 (left plot) and 0.67 (right plot).

correlation equals zero.

As the variance terms are in the denominator, the variables must have finite variances. In the case of heavy-tailed distributions the linear correlation coefficient does not exist (McNeil et al. (2005)) Therefore it is not recommended to use the linear correlation for analyzing the dependence of financial stocks. Non-elliptical distributions, especially in finance, demand alternatives for the linear correlation coefficient for empirical analysis. In chapter 4, Kendall's tau and realized dependence measures were used instead of Pearson's linear correlation.

2.2.2 Rank Correlation Coefficients

Two popular measures of association between two variables X and Y, Kendall's tau and Spearman's rho, deal with a form of dependence known as concordance. Both measures take into regard the order of elements in a random vector. This can be achieved with ranks.

Instead of the true values only the ranks of the variables, where R_i is a rank of X_i among $X_1, \ldots, X_n$, are regarded.

Basically, data can be ordered from the lowest to the highest value and enumerated. If two or more values of X (or of Y) are equal, e.g. X having the values $(5, 10, 10, 11)$, X is called to have ties. Then the values get the mean of the ranks they would have if they were different ($R = (1,\ 2.5,\ 2, 5,\ 4)^\top$). Yet, when assuming the variables X and Y to be continuous, ties occur with probability zero.

Ranks can also be thought of as the resulting numbers after using the probability transformation to get uniform marginals. A probability transformation of a random variable X with a continuous univariate distribution function F gives as a result a uniform distributed variable by computing $Y = F(X)$ with $Y \sim U(0, 1)$.

Rank correlation describes the dependence structure on these ranks. Spearman's rho and Kendall's tau are the two popular rank correlation coefficients and they use the numbers of concordant and discordant pairs.

Two observations (x_i, y_i) and (x_j, y_j) from a vector (X, Y) of continuous random variables are said

to be concordant if $x_i < x_j$ and $y_i < y_j$, or if $x_i > x_j$ and $y_i > y_j$. In contrast, two observations are discordant if $x_i < x_j$ and $y_i > y_j$, or if $x_i > x_j$ and $y_i < y_j$. Alternatively, concordance can be expressed as the case when $(x_i - x_j)(y_i - y_j) > 0$ and discordance with $(x_i - x_j)(y_i - y_j) < 0$. In the context of concordance, one looks at how two pairs of random variables "behave together".

Since ranks give the order of the data and the scale of the original values is not considered, rank correlation coefficient are a good tool to describe ordinal data as well as metric. Also, ranks are attractive for nonparametric procedures because they are not limited to a specific kind of distribution like the linear correlation coefficient.

In the following, the bivariate dependence measures Spearman's rho (denoted by ρ_S) and Kendall's tau (denoted by τ) are presented.

Spearman's Rho Named after Charles Spearman, the idea behind this coefficient ρ_S is to compute the correlation between the cumulative distribution functions $F_X(X)$ and $F_Y(Y)$ or similarly between the pairs (R_i, S_i) of ranks. In form of ranks, Spearman's rho can be defined for a sample of size n by

$$\rho_s = \frac{\sum_{i=1}^{n}(R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^{n}(R_i - \bar{R})^2 \sum_{i=1}^{n}(S_i - \bar{S})^2}} \in [-1, 1]$$

where $\bar{R}$ and $\bar{S}$ are the average ranks $\bar{R} = \frac{1}{n}\sum_{i=1}^{n} R_i$ and $\bar{S} = \frac{1}{n}\sum_{i=1}^{n} S_i$. Since both, R_i and S_i can be viewed as uniform distributed random variables, their mean is equal, $\bar{S} = \bar{R} = \frac{n+1}{2}$. When dividing the ranks by $(n+1)$, uniformly distributed variables in [0,1] are obtained. This leads to the support of the so-called empirical copula (see Genest and Favre (2007)).

When departing from the empirical sample to the general case, the formal definition in terms of the copula C associated with the joint distribution $F(X, Y)$ is given by

$$\rho_S(X, Y) = 3Q(C, \Pi) = 12 \int_0^1 \int_0^1 uv \, \mathrm{d}C(u, v) - 3 = 12 \int_0^1 \int_0^1 C(u, v) du \, dv - 3$$

According to Joe (1997), Spearman's rho and also Kendall's tau are invariant under strictly increasing transformations.

Kendall's tau Again based on ranks, another dependence measure is named after the British statistician Maurice George Kendall (1907-1983, Kendall and Gibbons (1990)). It considers the number of concordant pairs P_n and discordant pairs Q_n which are sample values for some sample size n. Using these quantities, the empirical version of Kendall's tau is defined as

$$\tau_n = \frac{P_n - Q_n}{0.5 \cdot n(n-1)}$$

When the X and Y are continuous random variables, then the case $(X_i - X_j)(Y_i - Y_j) = 0$ occurs with probability zero. So ties should not necessary be considered. The population version of Kendall's tau, for which τ_n is an asymptotically unbiased estimator (Genest and Favre (2007)), can be given by

$$\tau_{XY} = Q(C, C) = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1 = 4E(C(U, V)) - 1$$

From this definition it gets clear that Kendall's tau depends on the copula. Since the rank correlations and also the tail dependence coefficients are functions of the copula alone, they can be used in the parametrization of copulae (McNeil et al. (2005))

Kendall's tau can also be expressed using the probability function, for $d = 2$ this has the form

$$\tau = P\{(X_i - X_j)(Y_i - Y_j) > 0\} - P\{(X_i - X_j)(Y_i - Y_j) < 0\} = 2P\{(X_i - X_j)(Y_i - Y_j) > 0\} - 1$$

For elliptical distributions τ can easily be transformed to the linear correlation: $Corr(X, Y) = \sin(\frac{\pi}{2}\tau)$. A similar relation is found for Spearman's rho, $Corr(X, Y) = 2\sin(\frac{\pi}{6}\rho_S)$.

2.2.3 Coefficients of Tail Dependence

In the same manner as the previous dependence concepts, the tail dependence measures dependence non-parametrically and describes how large values of one random variable appear with large values of the other or equivalently for small values. However, tail dependence coefficients measure dependence between variables only in the upper-right and in the lower-left quadrant of a bivariate distribution. Similar to Kendall's tau and Spearman's rho, the tail dependence coefficients are invariant to increasing transformations.

Following the definition in Nelsen (2007), the upper and lower tail dependence parameters λ_U and λ_L are determined in the following way. For two continuous random variables X and Y with distribution functions F and G, respectively, the upper tail dependence parameter λ_U is the limit of the conditional probability that Y exceeds the α - quantile ($\alpha \in [0, 1]$) of G given that X is greater than the α - quantile of F as α approaches 1.

$$\lambda_U = \lim_{\alpha \to 1^-} P\left[Y > G^{-1}(\alpha) \mid X > F^{-1}(\alpha)\right]$$

This limit of the conditional distribution does not necessary exist. The analog definition for the lower tail dependence parameter is

$$\lambda_L = \lim_{\alpha \to 0^+} P\left[Y \leq G^{-1}(\alpha) \mid X \leq F^{-1}(\alpha)\right]$$

For copula applications this measure of dependence is very usefull, because the tail dependence parameters depend only on the copula and have relatively simple connection to C. If $X = (X_1, X_2)^\top$ is a continuous bivariate random vector with copula C, then the tail dependence coefficients can be calculated as

$$\lambda_U = \lim_{u \to 1^-} \frac{1 - 2u + C(u, u)}{1 - u}$$

$$\lambda_L = \lim_{u \to 0^+} \frac{C(u, u)}{u}$$

These tail dependence coefficients can take values in $[0, 1]$, whereas zero means independence. The normal distribution, for example has no upper and no lower tail dependence, that is $\lambda_U = 0$ and $\lambda_L = 0$.

Due to the connection between copulae and tail dependence, the gain from using copulae is the possibility to catch asymmetric dependence and tail-dependence (Hafner and Manner (2008))

Chapter 3

Copulae

Joint distribution functions of random vectors of risk factors contain enough information about the relation of the factors and their individual behavior. If one of the marginal distribution changes, the joint distribution function would also change. The concept of copulae is to isolate the information of the dependence structure from the marginal behavior. The dependence in form of the multivariate structure is summarized in a copula dependence parameter θ. One can distinguish between one-parameter and multi-parameter copulae. In this work, only one-dimensional copula dependence parameter are considered for the empirical application.

In the simple case when a multivariate normal distribution can be assumed, there are some standard procedures for estimation the joint distribution function. But for non-normal joint distributions the copula theory becomes very helpful. Copula models have an important advantage in practice - one can look beyond linear dependence structure and find more complex interconnections. McNeil et al. (2005) describe the copulae as an extremely useful concept and give a comprehensive definition. Yet, theoretical foundations of copulae are complex (Trivedi and Zimmer (2005)) and often simulation-based estimation is done. After some basic theory on copulae, such estimation procedures are presented in this chapter.

Definition. (Copulae)

A multivariate distribution function on the d-dimensional hypercube $[0,1]^d$ with standard uniform marginal distributions is called a copula with

$$\begin{aligned} C : [0,1]^d &\rightarrow [0,1] \\ \mathbf{u} = (u_1, \ldots, u_d) &\mapsto C(u_1, \ldots, u_d) \end{aligned}$$

satisfying

1. $C(u_1, \ldots, u_d) = 0$ if $u_i = 0$ for at least one $i = 1, \ldots, d$.
2. $C(u_1, \ldots, u_d) = u_i$ if $u_j = 1$ for all $j = 1, \ldots, d$ and $j \neq i$.
3. C is quasi-monotone on $[0,1]^d$.

For a parametric copula family $C(\mathbf{u}; \theta)$, the denotation includes some parameter $\theta \in \Theta \subset \mathbb{R}^p \, (p \geq 1)$, which differs in dimension and parameter space according to the family.

Some commonly used copulae are the Gaussian copula, Gumbel copula, Frank and Clayton copula. These characterize different dependence structures, the Clayton copula for example exhibits positive dependence and lower tail dependence whereas the Gumbel copula has positive dependence and upper tail dependence (Chen et al. (2006)). The Gaussian copula has symmetric positive and negative dependence but no tail dependence.

Sklar first introduced the notion of copulae in the form as above, as he flexibly decomposed the joint distribution function into the marginals and some function he called copula. (Lidan Großmaß (2013); Sklar (1959))

Theorem. (Sklar's theorem 1959)

Let F be a joint distribution function of a d-dimensional random vector $X = (X_1, \ldots, X_d) \in \mathbb{R}^d$ with the marginal distributions $F_1, \ldots, F_d$. Then there exists a copula $C : [0,1]^d \to [0,1]$ such that

$$F(x_1, \ldots, x_d) = C\{F_1(x_1), \ldots, F_d(x_d)\} \tag{3.1}$$

The copula C is unique if the marginal distributions F_i are continuous, for all $i = 1, \ldots, d$. In addition, if the copula C and the univariate distribution functions $F_1, \ldots, F_d$ are given, then a multivariate distribution function F with the marginals $F_1, \ldots, F_d$ can be constructed using formula (3.1). Further details can be found in McNeil et al. (2005).

From the relationship between copulae and multivariate distributions, as summarized in Sklar's theorem, a new multivariate distribution arises by changing some copula parameters or by replacing the marginal distribution functions with new ones. Considering the probability integral transformation $X_i = F_i(U_i)$, the formula (3.1) can be transformed in the following way

$$C(u_1, \ldots, u_d) = F\left\{F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\right\}, \quad u_1, \ldots, u_d \in [0,1]$$

where F_i^{-1} are the inverse marginal distribution functions, $i = 1, \ldots, d$. By this means the copula can directly be determined.

Beside the copula distribution function, the copula densities can be specified as usual densities of multivariate distribution. The density functions for the Clayton and the Gumbel copula are also needed for later estimation procedures.

Definition. (Copula density)

For a continuous copula C, the copula density is defined as

$$c(u_1, \ldots, u_d) = \frac{\partial^d C(u_1, \ldots, u_d)}{\partial u_1 \cdots \partial u_d}, \qquad u_1, \ldots, u_d \in [0,1]$$

Knowing the copula density, the univariate margins $F_1, \ldots, F_d$ and the corresponding univariate densities $f_1, \ldots, f_d$, the density of the the multivariate distribution F can be determined by

$$f(x_1, \ldots, x_d) = c\{F_1(x_1), \ldots, F_d(x_d)\} \prod_{i=1}^{d} f_i(x_i), \qquad x_1, \ldots, x_d \in \mathbb{R}$$

3.1 Copula Examples

For all copula functions an upper and lower bound can be found for all $\mathbf{u} = (u_1, \ldots, u_d)^\top \in [0,1]^d$ by applying the maximum and minimum function in the following way.

$$W(u_1, \ldots, u_d) \leq C(u_1, \ldots, u_d) \leq M(u_1, \ldots, u_d)$$

where $W(u_1, \ldots, u_d) = \min(u_1, \ldots, u_d)$ is called the lower Fréchet-Hoeffding bound and $M(u_1, \ldots, u_d) = \max\left(\sum_{i=1}^{d} u_i - d + 1, 0\right)$ the upper Fréchet-Hoeffding bound.

In the bivariate case, both the upper and lower bound are copulae and with $(u_1, u_2) \in [0, 1]^2$ these bounds have a simple form.

$$\max(u_1 + u_2 - 1, 0) \leq C(u_1, u_2) \leq \min(u_1, u_2)$$

For $d \geq 2$ only the upper bound M is a copula. Furthermore the Fréchet-Hoeffding bounds represent perfect negative (W) and perfect positive (M) dependence. This holds also for Kendall's tau and Spearman's rho, where W and M are reached for $\tau = 1 = \rho_S$ and $\tau = -1 = \rho_S$, respectively. This does not hold for the linear correlation coefficient (Joe (1997)).

Another extreme case is when the random variables are independent of each other. Then C is a product copula denoted by Π, which is simply the product of the random variables.

$$\Pi(u_1, \ldots, u_d) = u_1 \cdot \cdots \cdot u_d = \prod_{i=1}^{d} F_{X_i}(x_i)$$

This can be of course generalized for the n-dimensional case.

There are many functions satisfying the requirements for being a copula function. Two of the most popular copula classes are presented here.

3.1.1 Elliptical Copulae

The most popular distribution function, the Gaussian distribution, has some popular characteristic. It has a symmetric bell-shaped distribution. This symmetry is a main property of elliptical distribution. Elliptical distributions are an extension of multivariate normal distribution $\mathrm{N}_n(\mu, \Sigma)$, where μ is the mean and Σ is the covariance matrix. Gaussian and t copula are presented after a definition on elliptical distributions (cf. Franke et al. (2011)).

Definition. (Elliptical Distribution)

Let Y be a d-dimensional random vector, i.e. $Y = (Y_1, \ldots, Y_d)$. For some vector $\mu \in \mathbb{R}^d$, a non-negative definite symmetric $d \times d$ matrix Σ and some function $\phi : [0, \infty] \to \mathbb{R}^d$, the characteristic function $\varphi_{Y-\mu}$has the form $\varphi_{Y-\mu}(t) = \phi(t^\top \Sigma t)$ Then Y has an elliptical distribution with the parameters μ, Σ and ϕ.

Using Sklar's theorem elliptical copulae can be constructed from some specified elliptical distribution and marginal distributions. Two principal members of the elliptical copula are the Gaussian copula and the t-copula. They are presented briefly.

Gaussian Copula Let Φ denote the standard univariate normal distribution function and Φ_Σ the multivariate normal distribution function with correlation matrix Σ. Then the d-dimensional Gaussian copula can be constructed in the following way.

$$C_\Sigma(u_1, \ldots, u_d) = \Phi_\Sigma\left\{\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\right\}, \quad u_1, \ldots, u_d \in [0, 1]$$

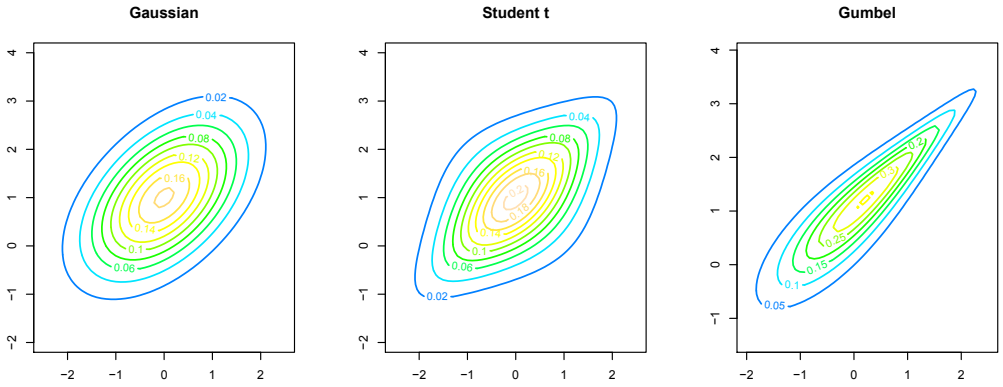

Figure 3.1.1: Contour plots of a Gaussian copula with correlation of 0.5, t - Copula with same correlation and 4 degrees of freedom and the Gumbel copula with dependence parameter equal 3. All three copulae are 2-dimensional.

The dependence parameter θ corresponds to the unknown correlation coefficients in Σ. For a bivariate normal distribution the linear correlation coefficient ρ can easily be transformed to Kendall's τ or to Spearman's ρ_S:

$$\begin{aligned} \tau &= \frac{2}{\pi}\arcsin(\rho) \\ \rho_S &= \frac{6}{\pi}\arcsin(\rho/2) \end{aligned}$$

The contour plot in figure 3.1.1. illustrates the joint symmetric dependence. Since the inverse of the normal distribution function is not easy to compute, the copula cannot explicitly be given. A special feature of the Gaussian copula is the implication of independence from zero correlation. If the correlation matrix Σ is the unit matrix I_d containing ones on the diagonal and zeros else, then the independence copula is obtained.

Student's t-Copula The t-distribution is an often used alternative for the normal distribution due to the property of heavier tails, that is, it has a higher tail index. Considering a standard univariate t-distributions, the cdf of a random variable$x \in \mathbb{R}^d$ is denoted by $t_\nu(x)$, where ν is the number of degrees of freedom.

Similar to the normal setting, the t-Copula can be constructed from univariate t - distribution functions and a joint t - distribution function $t_{\nu,\Psi}$ of a random vector $X \sim t_d(\nu, \mathbf{0}, \Psi)$ with correlation matrix Ψ (McNeil et al. (2005)),

$$C^t_{\nu,\Psi}(u_1, \ldots, u_d) = t_{\nu,\Psi}\left\{t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_d)\right\}$$

Here, the copula dependence parameter has two components: $\theta = (\theta_1, \theta_2) = (\nu, \Psi)$. The first component θ_1 describes the heaviness of the tails and as $\theta_1 \to \infty$ the t-copula corresponds to the Gaussian copula.

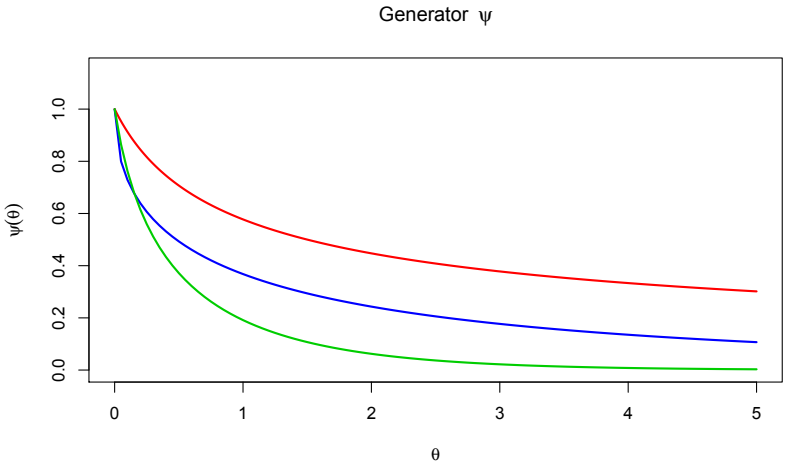

Figure 3.1.2: Generator function for the Clayton (red line), Gumbel (blue line) and Frank (green line) copulae, all with parameter $\theta = 2$.

3.1.2 Archimedean Copula Family

Archimedean copulae are an associative class of copulae. These dependence models are widely used in low-dimensional applications (McNeil et al. (2005)). In the case of elliptical copulae the inverse of the marginals is needed, so that the copulae don't have a simple closed form. In contrast to this, Archimedean copulae have an explicit formula due to their special structure. A d - dimensional Archimedean copula is a function $C : [0,1]^d \to [0,1]$ and it is defined by

$$C(u_1, \ldots, u_d) = \phi\left\{\phi^{-1}(u_1) + \phi^{-1}(u_2) + \cdots + \phi^{-1}(u_d)\right\}, \quad u_1, \ldots, u_d \in [0,1]$$

where $\phi : [0,\infty) \to [0,1]$ with $\phi(0) = 1$ and $\phi(\infty) = 0$ is called generator.

The Archimedean copula generator ϕ is a continuous, strictly decreasing and differentiable function which belongs to the class of Laplace transforms (Härdle and Okhrin (2009)). The derivatives must satisfy $(-1)^j \phi^{(j)} \geq 0, \quad j = 1, 2, \ldots$. In figure 3.1.2, three different generators are shown.

Only when ϕ and ϕ^{-1} have a closed form, the Archimedean copula is explicitly given. This is satisfied for many members of the Archimedean copula family, for example the four copulae in table 3.1. For the given Archimedean copulae the last column in table 3.1 provides the functional relationship f_τ between Kendall's tau and the generator function ϕ together with the copula parameter θ.

$$f_\tau(\theta) = 1 + 4 \int_0^1 \frac{\phi^{-1}(t,\theta)}{(\phi^{-1})'(t,\theta)} \mathrm{d}t \tag{3.2}$$

Using this mapping, an Archimedean copula model can be estimated from Kendall's tau.

Being members of the Archimedean copula family, the Gumbel and the Clayton copula are often used in practice, so they are presented in the following.

Gumbel Copula The Gumbel Copula is the only extreme value distribution in the group of archimedean copulae for the two dimensional case (Franke et al. (2011)). It exhibits upper tail dependence but does not allow for allow for lower tail dependence, akin to the Archimedean family of Joe. The generator function for the Gumbel copula is $\phi(x,\theta) = \exp\left\{-x^{1/\theta}\right\}$ with $1 \leq \theta < \infty$ and

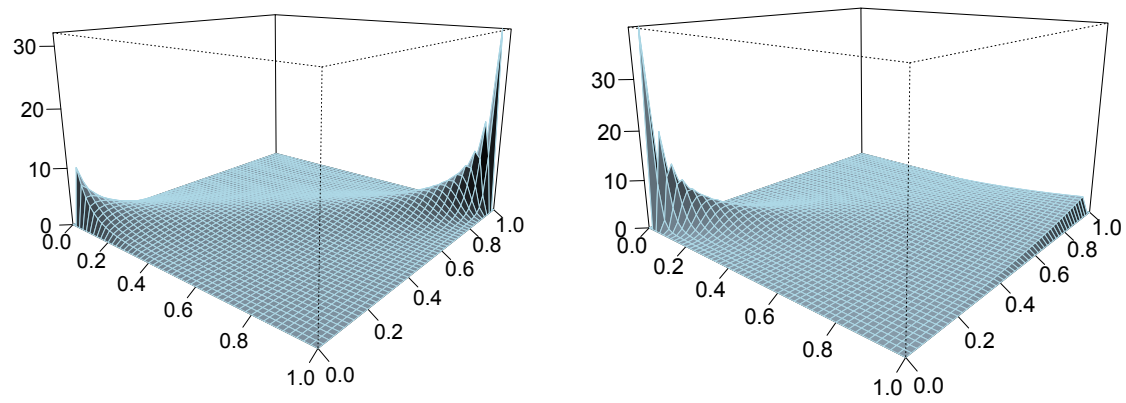

Figure 3.1.3: Density of 2-dimensional Gumbel (left) and Clayton copula, both with dependence parameter equal 3. The Gumbel copula generates upper right tail dependence while the Clayton copula has more mass on the lower tail.

$x \in [0, \infty)$. The copula function and its density for 2 dimensions are given by

$$\begin{aligned} C_\theta(u_1, \ldots, u_d) &= \exp\left[-\left\{ \sum_{j=1}^{d} (-\log u_j)^\theta \right\}^{1/\theta} \right] \\ c_\theta(u_1, u_2) &= \exp\left\{ -(\tilde{u}_1^\theta + \tilde{u}_2^\theta)^{1/\theta} \right\} (u_1 u_2)^{-1} \frac{(\tilde{u}_1 \tilde{u}_2)^{\theta-1}}{(\tilde{u}_1^\theta + \tilde{u}_2^\theta)^{2-1/\theta}} \left\{ (\tilde{u}_1^\theta + \tilde{u}_2^\theta)^{1/\theta} + \theta - 1 \right\} \end{aligned}$$

where $\tilde{u}_1 = -\log u_1$ and $\tilde{u}_2 = -\log u_2$.

When the parameter for the Gumbel copula takes its lower limit, $\theta = 1$, the product copula Π results. Whereas, when $\theta \to \infty$, the Gumbel copula take the Fréchet-Hoeffding upper bound M.

Clayton Copula Another famous member of the Archimedean family is the Clayton copula with the generator $\phi(x; \theta) = (1 + \theta x)^{-1/\theta}$ (alternative version in McNeil et al. (2005) $\phi(x; \theta) = (1 + x)^{-1/\theta}$) and its distribution function is given by

$$C_\theta(u_1, \ldots, u_d) = \left\{ \left(\sum_{j=1}^{d} u_j^{-\theta} \right) - d + 1 \right\}^{-1/\theta}$$

The dependence structure is captured by a positive parameter $\theta \in (0, \infty)$ (in a less strict case: $\theta \in [-1, \infty) \setminus \{0\}$) and the density can be determined for any dimension $d = 1, 2, \ldots$.

$$c_\theta(u_1, \ldots, u_d) = \prod_{j=1}^{d} \{1 + (j-1)\theta\} u_j^{-(\theta+1)} \left(\sum_{j=1}^{d} u_j^{-\theta} - d + 1 \right)^{-(\theta^{-1}+d)}$$

When the Clayton copula parameter θ takes the lower limit $\theta = -1$ the Fréchet-Hoeffding lower bound W results and the natural upper bound for $\theta \to \infty$ is the upper Fréchet-Hoeffding limit M meaning maximal dependence. The independence case is reached when $\theta \to 0$.

Copula	Generator $\phi(x,\theta)$	$\phi^{-1}(y)$	Parameter space
Independence	$-\log x$	e^{-y}	-
Gumbel	$\exp(-x^{1/\theta})$	$(-\log y)^{\theta}$	$\theta \in [1,\infty)$
Clayton	$(1+\theta y)^{-1/\theta}$	$\frac{1}{\theta}(x^{-\theta}-1)$	$\theta \in [-1,\infty),\ \theta \neq 0$
Frank	$-\frac{1}{\theta}\log\{1-(1-e^{-\theta})e^{-x}\}$	$\log\frac{\exp(\theta y)-1}{\exp(\theta)-1}$	$\theta \in [0,\infty)$

Table 3.1: Generator $\phi(t)$ and its inverse for some Archimedean Copulae with $x \in [0,\infty)$.

In contrast to the Gumbel copula, the Clayton copula does not allow for upper tail dependence. In figure 3.1.3 both copulae are shown and it can be seen that the Gumbel copula generates more probability mass on the right tails and the Clayton copula on the left tails. If one is interested in modeling lower and upper tail dependence, a mixture of the Gumbel (C^{Gu}) and the Clayton (C^{Cl}) copula can be constructed. A mixture of these two distributions is the distribution function $\lambda C^{Gu} + (1-\lambda)C^{Cl}$ where $\lambda \in (0,1)$ is a fixed real number Nelsen (2007).

3.2 Estimation of the Dependence Parameter

Copula estimation implies estimation of the parameters θ for the parametric copula belonging to a parametric family $\mathcal{C} = \{C_\theta, \theta \in \Theta\}$. There exist several approaches and they mainly differ in the assumption of an underlying distribution. For the multivariate copula-based models classical statistical inference theory is often not applicable, as pointed out by Joe (1997). Instead numerical methods replace the standard closed form estimators. Yet, the very powerful maximum likelihood (ML) theory can be applied for the parametric estimation, for which the models need specification of a distribution for the marginals and for the copula model.

There are several estimation methods for the copula parameter. Usually the estimation is performed by a fully parametric or stepwise parametric Maximum Likelihood method. The latter is called inference function for margins (IFM) method. Apart from that new methods are developed and active research is done.

3.2.1 Full Maximum Likelihood (FML)

Using Sklar's theorem, a copula C with parameter θ can be obtained when the a d-dimensional distribution function F and the univariate margins $F_1, F_2, \ldots, F_d$ with respective parameters $\alpha_1, \ldots, \alpha_d$ are available. Then the density of the distribution F can be represented as

$$f(x_1,\ldots,x_d;\alpha_1,\ldots,\alpha_d,\theta) = c\{F_1(x_1;\alpha_1),\ldots,F_d(x_d;\alpha_d),\theta\}\prod_{i=1}^{d} f_i(x_i;\alpha_i)$$

where c is the copula density defined by $c(u_1,\ldots,u_d) = \frac{\partial C(u_1,\ldots,u_d)}{\partial u_1 \partial u_2 \ldots \partial u_d}$. The maximum likelihood estimation (MLE) for the copula model requires this density function. Then the log-likelihood function of a random sample of independent and identically distributed (i.i.d.) vectors $x^{(j)} = (x_1^{(j)},\ldots,x_d^{(j)})^\top$

with T observations ($t = 1, \ldots, T$) can be decomposed into a copula part and a marginals part:

$$\begin{aligned} \log L(\alpha; x^{(1)}, \ldots, x^{(T)}) &= \sum_{t=1}^{T} \log f(x_1^{(t)}, \ldots, x_d^{(t)}; \alpha) \\ &= \sum_{t=1}^{T} \log\, c\left\{F_1(x_1^{(t)}; \delta_1), \ldots, F_d(x_d^{(t)}; \delta_d); \theta\right\} + \\ &\qquad \sum_{i=1}^{d} \sum_{t=1}^{T} \log f_i(x_i^{(t)}; \delta_i) \end{aligned} \tag{3.3}$$

where $\alpha = (\delta_1, \ldots, \delta_d, \theta)^\top$ is the full parameter vector containing the parameters from the marginals and the copula. The ML estimate of α is obtained by maximizing equation 3.3, $\hat{\alpha} = \arg\max_{\alpha} \log L(\alpha)$. Hence the maximum likelihood estimation can be performed simultaneously for the parameters of the margins δ_i and for the copula dependence parameter θ. This estimation method is also called full maximum likelihood. Some drawback of this algorithm is that when the dimension increases, the computation becomes too complex.

3.2.2 Inference for Margins (IFM)

The decomposed likelihood function in equation 3.3 suggests a separate estimation of the univariate parameters from the multivariate parameters. This leads to a two-stage estimation procedure, where the parameters of the margins are estimated

1. at the first stage by maximizing the log likelihood
   $\log L_i(\delta_i) = \sum_{t=1}^{T} \log f_i(x_i^{(t)}; \delta_i)$ with $i = 1, \ldots, d$
   with respect to parameter of the margin, δ_i

2. and in the second stage regarding the pseudo log-likelihood function

$$\log L(\theta, \hat{\delta}_1, \ldots, \hat{\delta}_d) = \sum_{t=1}^{T} \log\, c\left\{F_1(x_1^{(t)}; \hat{\delta}_1), \ldots, F_d(x_d^{(t)}; \hat{\delta}_d), \theta\right\}$$

   with the estimated dependence parameter $\hat{\theta} = \arg\max_{\theta} \log L(\theta, \hat{\delta}_1, \ldots, \hat{\delta}_d)$

This procedure was recommended in Joe (1997) and it is called inference for margins. In contrast to the full maximum likelihood estimation, the numeric complexity is drastically reduced as explained in Franke et al. (2011).

3.2.3 Semi-Parametric Estimation

Alternatively, the parametric marginal distributions can be replaced non-parametrically by the empirical distribution functions F_n. If the copula comes from a parametric family this would lead to a semi-parametric MLE. Since the copula likelihood is required, it is not appropriate for high-dimensional copula models where the copula density is not easily obtained.

For the case, when the likelihood of the copula is either not known in closed form, or is complicated to obtain and maximize there exist alternatives to the maximum-likelihood estimation. As a standard tool for non-parametric inference, rank correlation coefficients can be regarded. If the copula belongs to a specific family, like the Archimedean family, there exist formulas to obtain the copula parameter

Copula	$\theta = f_\tau^{-1}(\tau)$	$\tau = f_\tau(\theta)$
Gaussian	$\theta = \rho = \sin(\tau\frac{\pi}{2})$	$\frac{2}{\pi}\arcsin(\rho)$
Gumbel	$\theta = \frac{1}{1-\tau}$	$1 - \frac{1}{\theta}$
Clayton	$\theta = \frac{2\tau}{1-\tau}$	$\frac{\theta}{\theta+2}$

Table 3.2: Relationship between the copula dependence parameter θ and Kendall's τ.

using Kendall's tau. That is, in some cases, the mapping from the parameter(s) of the copula to dependence measures like Spearman's ρ or Kendall's τ is known in closed form, thus allowing for method of moments (MM) or generalized method of moments (GMM) (Oh and Patton (2013)). In general, this mapping, e.g. $f(\tau) = \theta$, is not known.

In table 3.2, the relation between the copula dependence parameter θ and Kendall's τ is given for three copulae. Kendall's tau, just as Spearman's rank correlation and quantile dependence are functions only of the copula and according to Nelsen (2007), for a pair of continuous random variables (X_i, X_j) with univariate distributions F_i and F_j, these measures are defined in chapter 2 with $u = F_i(X_i)$ and $v = F_j(X_j)$.

In practice, it can be assumed that the underlying stock returns X_i with $i = 1, \ldots, d$ have a normal distribution. The components from the estimated correlation matrix can be used to get values of Kendall's tau by the transformation $\tau_{ij,t} = \frac{2}{\pi}\arcsin\rho_{ij,t}$, where $\rho_{ij,t}$ is the linear correlation coefficient and $\tau_{ij,t}$ Kendall's rank coefficient between X_i and X_j for day t. A subsequent transformation of Kendall's tau to the copula parameter θ can be done with the following mapping, if the copula belongs to the Archimedean family with generator ϕ,

$$f_\tau(\theta) = 4\int_0^1 \phi_\theta^{-1}(\nu)/(\phi_\theta^{-1})'(\nu)d\nu + 1$$

This leads to an explicit and invertible relationship between Kendall's tau and the respective dependence parameter. For a dimension $d > 2$ the following step needs to be done.

$$\hat{\theta}_t^{Ad\,hoc} = \frac{2}{d(d-1)}\sum_{i<j} f_\tau^{-1}(\hat{\tau}_{i,j,t}^G)$$

This estimation method will be illustrated in a copula model of five equity returns in chapter 4. Also, other estimation method have been considered for copula-based multivariate models:

Canonical Maximum Likelihood (CML) The Canonical Maximum Likelihood estimation is a semi-parametric copula estimation method also called pseudo-likelihood. In the following it is related to the full maximum likelihood estimation in the bivariate case.

Given a time series of prices S_t, the log-returns are computed as $X_t = \log(S_t/S_{t-1})$. The log-return process $\{X_t\}$ can be modeled as

$$X_{j,t} = \mu_{j,t} + \sigma_{j,t}\eta_{j,t}$$

where $\sigma_{j,t}^2 = E\left\{(X_{j,t} - \mu_{j,t})^2|\mathcal{F}_{t-1}\right\}$ is the conditional variance of $X_{j,t}$ given $\mathcal{F}_{t-1}$ and

$\eta_t = (\eta_{1,t}, \ldots, \eta_{d,t})^\top$ are the standardized innovations for $j = 1, 2, \ldots d$. By this, the time series X_t can be decomposed into two parts, the predictable conditional mean $\mu_{j,t} = E(X_{j,t}|\mathcal{F}_{t-1})$ and the unpredictable component, $\varepsilon_{j,t} = \sigma_{j,t}\eta_{j,t}$, where $\mathcal{F}_t$ is the information available at time t. The innovations η_t are independent of $\mathcal{F}_t$ and i.i.d. with $E(\eta_{j,t}) = 0$ and $E(\eta_{j,t}^2) = 1$ for all $j = 1, \ldots, d$

(Chen et al. (2006)). Furthermore, the multivariate innovation η_t has a distribution function F_η with

$$\eta \sim iid\, F_\eta = C(F_1(\eta_1), \ldots, F_d(\eta_d); \theta_0)$$

where F_i are the unknown, true continuous marginals of the innovations η_i, $i = 1, \ldots, d$ and C is the belonging copula function with the unknown copula parameter θ_0.

In the maximum likelihood approach both the copula density function and the single marginals are regarded for the maximization. For two random variables the standardized residuals η_a and η_b are obtained having the univariate density functions f_1 and f_2 and distribution functions F_1 and F_2, respectively. Then the maximum likelihood estimation would be

$$\hat{\theta} = \max_{\theta} \sum_{t=1}^{n} \{\log f_1(\eta_{a,t}) + \log f_2(\eta_{b,t}) + \log\, c(F_1(\eta_{a,t}), F_2(\eta_{b,t}), \theta)\} \tag{3.4}$$

However, the Canonical Maximum Likelihood (CML) method, proposed by Genest et al. (1995), is a semi-parametric estimation procedure in which no assumptions are made about the parametric form of the marginal distributions. Empirical counterparts are substituted for the marginals. Then the copula parameter is estimated via maximum likelihood. Therefore the expression in equation(3.4) is shortened as

$$\hat{\theta}^{CML} = \max_{\theta} \sum_{t=1}^{n} \log c(F_{1n}(\eta_{a,t}), F_{2n}(\eta_{b,t}); \theta)$$

where $F_{in} = n/(n+1)\hat{F}_{in}$ with $i = 1, 2$. Of course, this can be extended to higher dimensions d.

3.3 Realized Copula

A realized copula results when realized dependence measures like the realized covariance, estimated from high-frequency data are used for constructing the copula function. This concept was brought forward by Fengler and Okhrin (2012). They used realized kernels as an estimator for covariation and applied Hoeffding's lemma on the data and the copula model to estimate the copula parameter.

Lemma. *(Hoeffding 1940)*

Let X_1 and X_2 be two random variables with the marginal distributions F_1 and F_2, respectively, and let F be their joint distribution, such that $E(|X_1|)$, $E(|X_2|)$and $E(|X_1X_2|)$ are all finite.

$$Cov(X_1, X_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \{F(x_1, x_2) - F_1(x_1)F_2(x_2)\} dx_1 dx_2$$

This Hoeffding's identity makes it possible to express the covariation of X_1 and X_2 in terms of their marginal and joint distributions (Joe (1997), Nelsen (2007)). This lemma can be used for the copula setting, when the joint distribution is replaced by the copula function,

$$\sigma_{ij}(\theta) = \int \int [C_\theta\{F_i(x_i), F_j(x_j)\} - F_i(x_i)F_j(x_j)] dx_i dx_j \tag{3.5}$$

For the multivariate normal distribution this relation simplifies to $\sigma_{ij} = \theta$. However, for the general case some estimate is found for the daily covariation σ_{ij} and the right part of equation 3.5 is estimated by using numerical integration. An estimate for θ is found, such that the equation is approximately fulfilled.

In detail, for each day some estimate $h_{ij,t}$ with elements of a measured covariance matrix is taken for $\sigma_{ij}(\theta)$ and the right part of equation 3.5 is abbreviated with $f_{ij}(\theta_t)$. By forming the difference $g_{ij}(\theta) = h_{ij,t} - f_{ij}(\theta)$ for d random variables with $i, j = 1, \ldots, d$ and $i < j$ an estimate for the copula parameter can be found by minimizing $\hat{\theta}_t^{MM} = \arg\min_{\theta} g^T(\theta) W g(\theta)$. Here, W is a positive definite weight matrix. For the empirical analysis the n-dimensional unit matrix $\mathbf{I}_n$ with $n = d(d-1)/2$ was chosen for this weight matrix. Fengler and Okhrin (2012) highlighted that this approach is in a method-of-moments type of fashion.

Chapter 4

Empirical Data Analysis

For this work high frequency intraday data was taken from Lobster, an online limit order book data tool that provides data of NASDAQ traded stocks[1]. For five stock, Oracle (ORCL), IBM, Pfizer (PFE), Google (GOOG) and Exxon (XOM), a time series consisting of trade data over the time period 01.10.2010 - 21.06.2013, a total of 684 days, was available.

In addition to this, daily price data was taken from Yahoo for the same stocks but for a longer period, 07.10.2008 to 21.06.2013. By this way, daily data contains 500 more observations. Besides, only the adjusted prices were considered.

The intraday data from Lobster included two csv-files for each day which are called message file and orderbook file. The message file contains six columns: **time** (in seconds from midnight), **type**, **order id**, **size**, **price** (dollar price times 10000) and **direction** (-1 for sell and 1 for buy limit order). The **time** and the **price** column are the most important ones, others like **type** where used for cleaning procedures.

time	type	order id	size	price	direction
34200.33789681	5	6165846	300	1352900	-1
34200.503721437	1	1841681	250	1355900	-1
34200.581503807	1	6332816	100	1352800	1
...	...	...	...	...	...

Table 4.1: High frequency intraday data taken from Lobster in message file. First three rows of IBM data for 2010-10-01.

The orderbook file was only used for cleaning purposes, when observations were deleted due to misrecording, for example. Depending on the number of price levels, the orderbook file contains columns for ask and bid prices and ask and bid sizes of minimum level 1 up to a selected level.

A very important step before working with the data, is to clean and filter the data set. The cleaning steps are explained in the next section.

For the cleaning steps, the synchronization and the empirical analysis, R was used which is a programming language and environment for statistical computing.

[1] LOBSTER: Limit Order Book System - The Efficient Reconstructor at Humboldt Universität zu Berlin, Germany. http://LOBSTER.wiwi.hu-berlin.de

4.1 Data Preparation

For the application of realized copulae a 5-dimensional log price process $Y = (Y_1, \ldots, Y_5)^\top$ was regarded. Some descriptive statistics can be found for the daily price data in table 4.2 and particular for the log-returns in table 4.3.

	Minimum	Median	Mean	Maximum	Std. Deviation	Skewness	Kurtosis
ORCL	24.21	30.71	30.50	36.07	2.61	-0.06	-0.77
IBM	127.74	184.80	177.62	212.58	21.60	-0.54	-0.80
PFE	14.73	20.19	20.86	30.34	4.04	-0.73	-0.73
GOOG	474.88	612.43	640.32	915.89	97.66	0.94	0.15
XOM	57.31	80.00	79.15	91.54	7.69	-0.29	-0.29

Table 4.2: Descriptive analysis of daily prices (Yahoo)

	Minimum	Median	Mean	Maximum	Std. Deviation	Skewness	Kurtosis
ORCL	-0.12	0	0	0.06	0.02	-1.25	7.13
IBM	-0.09	0	0	0.06	0.01	-0.67	6.09
PFE	-0.05	0	0	0.06	0.01	-0.09	2.66
GOOG	-0.09	0	0	0.12	0.02	0.32	10.81
XOM	-0.01	0	0	0.05	0.01	-0.32	3.13

Table 4.3: Descriptive analysis of daily log-returns (Yahoo)

From table 4.3 it can be seen that most of the assets have a negative skewness of their log-returns. That indicates fat left tails. Following the definition in Franke et al. (2011), the skewness of a random variable X is defined as

$$S(X) = \frac{E\left[(X-\mu)^3\right]}{\sigma^3}$$

where μ is the mean and σ^2 is the variance of X.

Furthermore, the kurtosis of all asset returns, except Pfizer, exhibit higher values then 3, which is the kurtosis value for normal distributed random variables. Therefore the empirical univariate distribution of the returns is leptokurtic, which means that there are many positive and negative outliers but also many values around the center (Franke et al. (2011)). The univariate distributions of all five asset returns can be seen in figure 4.1.1.

Since the available data is real market data, one has to deal with non-synchronous trading and irregularly spaced observations over the time interval $[0, T]$ where T denotes the number of observed days, i.e. $T = 684$.

As the first step, the cleaning procedures proposed by Barndorff-Nielsen et al. (2009) were applied. These are summarized in the table 4.5.

4.1.1 Cleaning Procedures Proposed by Barndorff-Nielsen

Price data was observed within the day between 9.30 am and 4 pm. Observations out of these bounds should be deleted. This rule is denoted by P1 in Barndorff-Nielsen et al. (2009). If the price (ask or bid price) equals zero, then this observation is canceled (P2). For trade data, which was used in this work, only entries with normal sale condition must be used (T2). The data obtained from Lobster has 7 types of transaction (compare table 4.4) where only executions are regarded as normal sale condition.

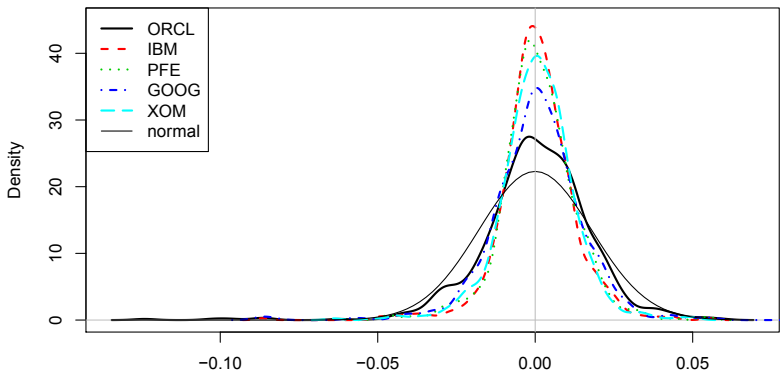

Figure 4.1.1: Kernel density estimates for the daily log-returns of ORCL, IBM, PFE, GOOG and XOM. The normal density (thin black line) has mean=0 and σ equal the sample standard deviation of ORCL.

Type	Description
1	Submitted new order
2	Cancellation (partial deletion of order)
3	Deletion
4	Execution against visible oder
5	Execution against hidden order
7	Halt

Table 4.4: Types of transaction for data from Lobster

Due to this step, the daily data size reduced enormously. In step T3, again a lot of entries were deleted because there were many observations at the same instant, meaning per second. In the message file, time is given in a nanosecond precision, as can be seen from table 4.1, and often times of subsequent observations differ only in nanoseconds with same prices. The time was rounded to seconds and multiple observations were replaced by the median price.

At last, from the ask p_A and bid prices p_B the bid-ask spread $|p_B - p_A|$ was computed. Only entries with prices within the bounds $p_B - |p_B - p_A| \leq p \leq p_A + |p_B - p_A|$ were kept. For this rule, T4, and all the other stated above, the data reduction is summarized in table 4.5.

	n	P1	P2	T2	T3	T4	n^{new}
Oracle	356749	0	0	329489.8 (92.4%)	23555.8 (6.6%)	0	33024.7 (0.92%)
IBM	72380	0	0	62826.7 (86.8%)	6699.2 (9.3%)	1	2965.1 (4.10%)
PFE	217896	0	1	202131.2 (92.8%)	13198.9 (6.1%)	0	1977.2 (0.91%)
GOOG	65822	0	1	53547.9 (81.4%)	9052.0 (13.8%)	3.33	3254.8 (4.94%)
XOM	350284	0	2	327093.0 (93.4%)	18626.0 (5.3%)	29	4432.0 (1.27%)

Table 4.5: Data reduction due to cleaning procedures. Daily averages of raw data size n and final data size n^{new}, daily average numbers of deleted entries in P1, P2, T2, T3 and T4 related to all cases where cleaning was necessary; percentages of data reduction in brackets (related to n).

Since this work looks at the interdependence of assets, the seasonality doesn't need to be removed. On the contrary it gets more interesting how the assets move together in special periods. Also, in Lidan Großmaß (2013) it was found that estimates of the copula dependence parameter using seasonal adjustment are close to those without adjustment.

For each day the 5 assets differ in the number of observations and also for the time points when each price was observed. After Cleaning the data, the data loss is between 96 % and 98 %. Altogether the worst loss of data was for Oracle, Pfizer and Exxon with about 99%. This could be improved either by choosing only assets with equally high recording frequencies or by some blocking strategy, as proposed by Hautsch et al. (2009).

4.1.2 Synchronizing

One of the major challenges of high frequency financial data is the fact that data is irregularly spaced and non-synchronous. This makes it hard to find observations for each stock of a portfolio for a specific point in time. It gets even worse, if one asset is almost not traded, so that a combination with a frequently traded asset would cost much of the data from the latter asset. Barndorff-Nielsen et al. (2009) pointed out that non-synchronous trading is one cause of bias and needs certain treatment like their proposed refresh time method.

For treating this problem the procedure of refreshed times, as defined in Barndorff-Nielsen et al. (2008b), is applied, adopting their denotation for this procedure. The goal in refresh time sampling is to find points in time for which prices can be found from all assets and therefore generate a common time vector. This setting is illustrated schematically in figure 4.1.2. For each day, this new time vector τ, called refresh time are computed.

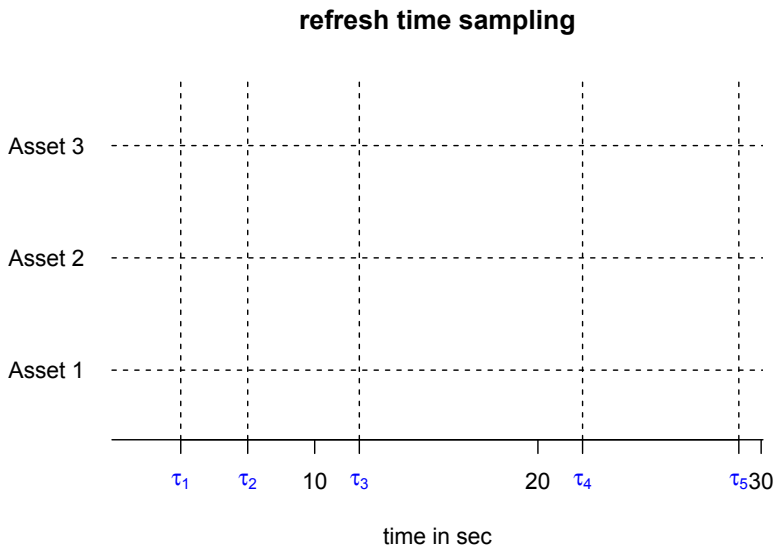

Figure 4.1.2: Refresh time sampling scheme. The refresh times are indicated in blue on the x-axis.

Definition (Refresh time) For d assets the first refresh time is defined as $\tau_1 = \max(t_1^{(1)}, \ldots, t_1^{(d)})$. All subsequent values for the refresh time are computed using the formula

$$\tau_{j+1} = \max \left\{ t^{(1)}_{N^{(1)}_{\tau_j}+1}, \ldots, t^{(d)}_{N^{(d)}_{\tau_j}+1} \right\}$$

where the superscript stands for the asset and the subscript for the position in the within-day time vector, e.g. $t_2^{(1)}$ is the second observed time for the first asset. In the beginning, every asset has a different number of observations and $N_s^{(i)}$ denotes the position in the time vector $t^{(i)}$ of asset i that corresponds to the time s within some fixed day.

By applying this refresh time sampling, the times of the multivariate data set are synchronized. Of course, the least frequently traded asset determines the final number of refreshed times.

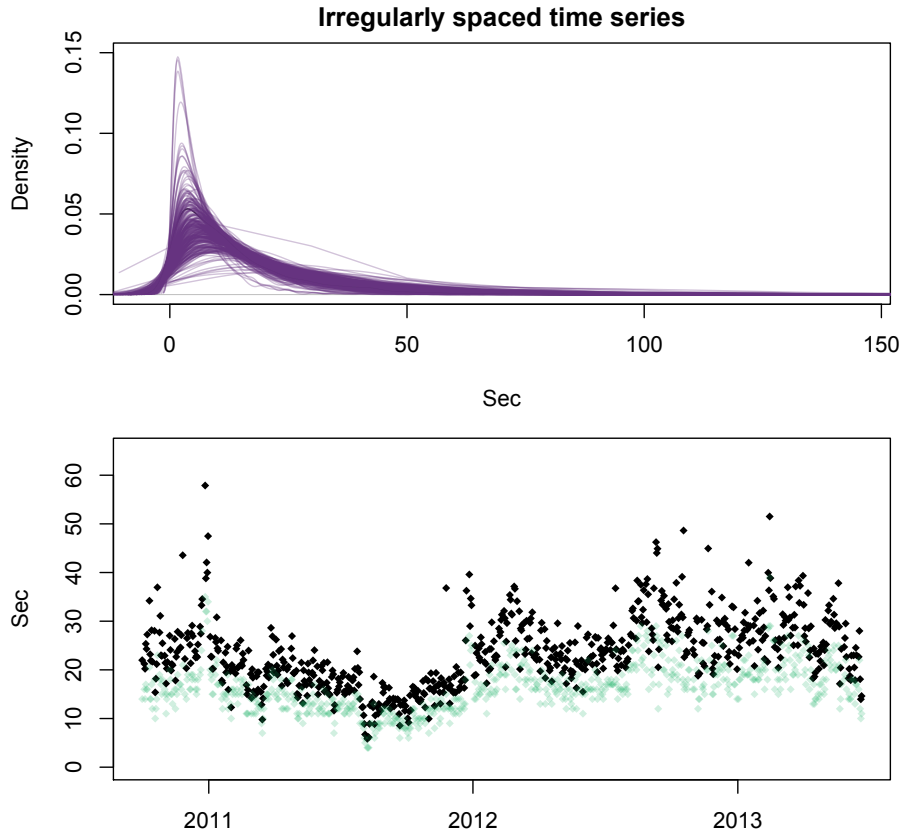

Figure 4.1.3: Daily distributions (upper) and daily averages (lower panel) of time interval length between observations for all 684 days. Irregularly spaced price observations (after data cleaning and synchronization) are mostly observed every second, but alltogether the mean intervall length is 23.4 seconds. In the lower panel, daily mean (black points) and daily median (green points) intervall lengths are plotted.

Summarizing the data preparation steps, in the start every stock had a different amount of available data for each day and during the day the observations were irregularly spaced. The number of observations for an asset for one day fluctuated between $16,660$ and $1,769,000$. The stocks XOM and ORCL had the largest numbers of observations, all together 243 and 239 millions for all 684 days, respectively.

After applying filtering rules, refreshing time and synchronizing all five stocks prices the overall amount of data is $748,062$ observations which corresponds to an daily average of 1093.66 observations

(minimum 221, median 1021 and maximum 4009 observations).

In figure 4.1.3, it can be seen that the final data set contains irregularly spaced observations in time, yet this available data will be called tick data since the prices are mostly observed every second.

Already in 2001 it was found by Andersen et al. (2001) that tick-by-tick prices, that means observations taken every second, are rarely available at equally-spaced time points. Some form of interpolation or other methods are needed to get an evenly-spaced time series. Due to market microstructure noise it is advisable to work with a lower sampling frequency, like 5 - minute returns. For this reason, all analysis was also made for 5 - minute returns. The construction of 5 - minute returns will be detailed later.

In section 4.2, an interesting results were found regarding different sensitivity of realized kernel and realized volatility estimates to changing sampling frequency from seconds to minutes.

The next step is to use the final data set to obtain log-returns.

The n'th intraday return within a day can be get as the change of the log prices during the corresponding period between two observations:

$$r_{t,n} = \log(p_{t,n}) - \log(p_{t,n-1}) \quad \text{with} \quad t = 1, \ldots, T,\ n = 1, \ldots, N_t$$

where $T = 684$, p_t is the price vector for day t and N_t is the number of intraday observations for day t.

In an equally weighted portfolio consisting of the five stocks ORCL, IBM, PFE, GOOG and XOM all log-returns are combined for all observed days and the overall return movement can be seen in figure 4.1.4. Until around July or August 2011 the returns movement was more calm. In August 2011, however, a type of crisis happened in the USA due to Standard & Poor's downgrading of the US credit rating from AAA to AA-plus (Brandimarte and Bases (2011)).

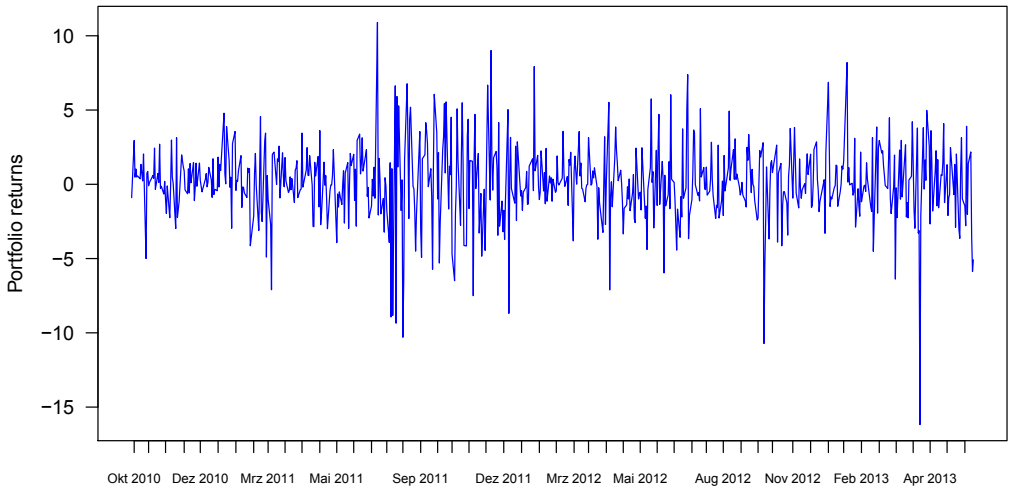

Figure 4.1.4: Returns of a linear portfolio of the five regarded stocks

4.2 Estimation of Covariation

In practice, the true log prices Y_t are latent random variables and only noisy observations can be made, that is $\widetilde{Y}_t = Y_t + U_t$ where U_t is some noise factor and Y_t is the efficient price. Liquidity effects, discrete price grids, bid-ask bounce effects or mis-recording can be sources for the noise U_t, which is the market microstructure noise. Due to this, the variation in intraday returns can be attributed to the efficient price and to the microstructure noise. Particular for very high sampling frequency, this noise gets severe (Duan et al. (2011)).

For estimating ex-post variation of asset prices, the quadratic variation ofY_t is regarded. According process theory can be found in Jin and Maheu (2013) and Barndorff-Nielsen et al. (2008a) where the quadratic variation of Y can be given as

$$[Y](1) = \operatorname*{plim}_{n\to\infty} \sum_{j=1}^{n} \{Y(t_j) - Y(t_{j-1})\} \{Y(t_j) - Y(t_{j-1})\}^\top$$

with an assumed time interval $[0, 1]$. Based on the the theory of quadratic variation, two volatility models, realized volatility and realized kernels, based on high-frequency intraday returns are constructed.

4.2.1 Realized Volatility

The realized volatility is a non-parametric estimator of volatility and it is defined as the sum of squared intraday returns,

$$RV_t = \sum_{s=1}^{n_t} r_s^2$$

where n_t is the number of returns during a trading day t. The realized covariance estimator can be constructed as

$$RCov_t = \sum_{i=1}^{n_t} r_{t,i} r_{t,i}^\top$$

where $r_{t,i}$ is the ith intraday return vector (Jin and Maheu (2013)). These measures of covariance matrices will be termed RV in the next sections.

If the observed high-frequency prices were not affected by the market microstructure effect U_t, the realized volatility would converge to the quadratic variation of the price process as the sampling frequency goes to infinity (Chen et al. (2010) or Andersen and Bollerslev (1998)). Then there would not be the search for alternative, efficient estimators for quadratic variation and the realized kernels would not be covered much here. But there is clear evidence for market microstructure noise as found in the following and the simple realized variance does not give satisfying results. The realized kernel estimator is reported to perform well in presence of market microstructure effects in Barndorff-Nielsen et al. (2008a).

Some direct way to reduce microstructure effects is to construct realized volatility based on a lower frequency of observations. This was done by looking only on five-minute returns instead of one-second returns.

Furthermore, estimation of volatility has a strong influence on the accuracy of the VaR measurement. The key is an accurate measure of the covariance matrix, based on intraday high-frequency return data. One kind of volatility models is the historical volatility model that is based on historical return data with daily periods with Realized Volatility (RV) being one important type of volatility

model. Contribution to this statistical area can be found in Taylor and Xu (1997) and Andersen and Bollerslev (1998).

4.2.2 Realized Kernel

The second approach, which is more robust to market microstructure effects, is the realized kernel (RK) estimator, which was proposed by Barndorff-Nielsen et al. (2008a). In Barndorff-Nielsen et al. (2009) the strength of realized kernels in presence of market microstructure noise was shown as the realized kernel gave similar results for trade and quote data. These data types differ in the sampling frequencies. Therefore the microstructure effect would manifest itself in different estimation results for these different types of financial databases. However, it was stressed, that data cleaning is a necessary step before estimation.

The formal definition of realized kernels for day t is

$$K(X) = \sum_{h=-H}^{H} k\left(\frac{|h|}{H+1}\right)\gamma_h, \qquad \text{with } \gamma_{t,h} = \sum_{j=|h|+1}^{n_t} r_{t,j} r_{t,j-h}^{\top} \tag{4.1}$$

where $\gamma_{t,h}$ is the hth realized autocovariance for day t and $k(h)$ is a weight function. Note, that $\gamma_{t,0}$ is the realized variance. As input, X is the data matrix of log returns for one day. The realized autocovariance sums up $d \times d$ matrices computed as the Cartesian product of the return vector $r_{t,j} = (r_{t,j,1}, \ldots, r_{t,j,d})^{\top}$ with its h- lag, $r_{t,j-h} = (r_{t,j-h,1}, \ldots, r_{t,j-h,d})^{\top}$.

As supported by Barndorff-Nielsen et al. (2009) the Parzen kernel was taken here for the weight function and it is defined as

$$k(t) = \begin{cases} 1 - 6t^2 + 6t^3 & 0 \leq t \leq 0.5 \\ 2(1-t)^3 & 0.5 \leq t \leq 1 \\ 0 & t > 1 \end{cases}$$

Since the Parzen kernel is in some way a symmetric smoother of the autocovariances, it is shown from -1 to 1 in figure 4.2.1. The negative values of h in formula 4.1 are taken in as absolute values in the weight function, therefore the parts of the autocorrelations that can be described as leads and those as lags are weighted equally. By using the Parzen kernel, the realized kernel give a non-negative estimate of the covariance matrix (see Barndorff-Nielsen et al. (2011) for further details). Other weighting kernels have worse efficiency (e.g. cubic kernels in Barndorff-Nielsen et al. (2008a)) or consistency (Bartlett kernel).

The bandwidth computation was done as proposed in Barndorff-Nielsen et al. (2009).

$$H = \xi^{4/5} c^* n^{3/5}$$

where n is the data size after refreshing time, c^* is a kernel weight specific constant (for the Parzen kernel $c^* = 3.5134$) and ξ depends on the unknown integrated quarticity. Simply said, ξ is the combination of the realized variance estimator based on high frequency returns involving microstructure and lower frequency of 20-minute returns. This bandwidth is a good trade-off between asymptotic bias and variance, that is when the sampling frequency is too high or too low, respectively. Some examples together with an overall average of the computed bandwidths are presented in table 4.6. The bandwidth parameter H controls the number of leads and lags used for all returns in a day.

After cleaning and synchronizing the high-frequency vector of returns, estimates for the daily covariance matrix $RCOV_t$ are computed using RV and RK. The diagonal elements are the realized

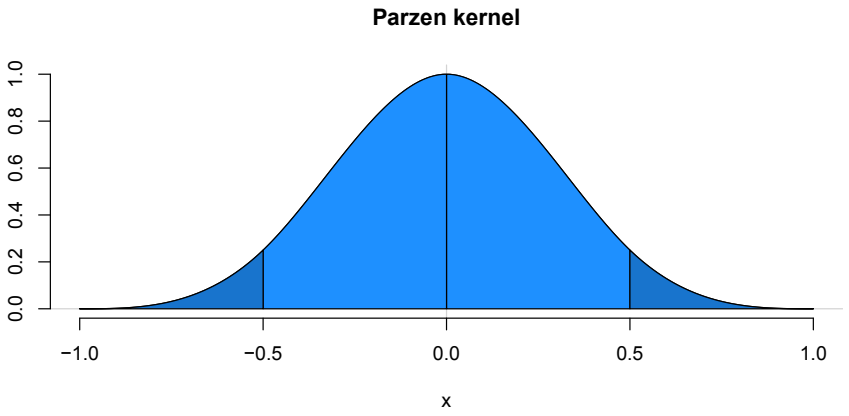

Figure 4.2.1: Parzen Kernel satisfies the necessary smoothness conditions, $k'(0) = k'(1) = 0$.

variances that are an ex post measure of the variance for the respective asset. The realized correlation between asset i and j is also obtained by $RCOV_{t,ij}/\sqrt{RCOV_{t,ii}RCOV_{t,jj}}$ where $RCOV_{t,ij}$ is the element from the i-th row and j-th column.

	01.10.2010	04.10.2010	05.10.2010	Average over all days
ORCL	15.15	12.63	12.65	14.41
IBM	13.79	13.73	14.01	14.36
PFE	15.21	15.84	12.62	14.30
GOOG	13.96	13.63	14.86	14.30
XOM	15.07	12.79	15.04	14.28
Ø	15	14	14	14.82

Table 4.6: Estimated Bandwidths for Realized Kernel, for 3 days and an average value for all 684 days

4.3 Regular Spaced Observations

Additionally to the irregular spaced tick data, 5 - minutes returns were also used. By this the condition of regular spaced observations was provided and also the market microstructure effect should not be large at this frequency. For computing 5 - minute returns, in the most cases there were enough observations near the 5 - minute grid points. For periods like the end of the day, sometimes there were too few observations, so the last chosen point in the 5 - minute grid was taken for the next 5 - minute return.

		High frequency irregular spaced data					5 minute returns				
	01.10.2010	ORCL	IBM	PFE	GOOG	XOM	ORCL	IBM	PFE	GOOG	XOM
RK	ORCL	1	0.628	0.435	0.499	**0.438**	1	0.627	0.316	0.317	0.507
	IBM		1	0.392	0.440	0.390		1	0.290	0.380	0.580
	PFE			1	0.436	0.439			1	0.476	0.536
	GOOG				1	0.390				1	0.544
	XOM					1					1
	01.10.2010	ORCL	IBM	PFE	GOOG	XOM	ORCL	IBM	PFE	GOOG	XOM
RV	ORCL	1	0.322	0.160	0.298	**0.313**	1	0.571	0.359	0.339	0.501
	IBM		1	0.130	0.248	0.345		1	0.302	0.400	0.567
	PFE			1	0.147	0.207			1	0.489	0.509
	GOOG				1	0.218				1	0.514
	XOM					1					1

Table 4.7: Correlation estimates based on realized kernel and realized covariance on 01.10.2010.

For both, the irregular spaced time series and the 5 - minute time series of returns all estimates were computed. In figure 4.7, the estimates for RK and RV are given for the day 01.10.2010 as an example. For tick data there is as a clear difference between RK and RV estimates but for 5 - minute returns they equal each other more.

Under the assumption that the time interval is fixed, the ex-post realized volatility is an unbiased volatility estimator (Corsi (2009)). From table 4.7, this gets obvious, because for 5 - minute returns the estimates from RK and RV approach. Apparently, the simple realized variance is biased downwards with a too high sampling frequency of seconds with irregular spaced intervals.

It was interesting to see how the realized kernel estimates for tick data differ from the respective estimates based on 5 - minutes returns. The same was done with the RV estimates. From table 4.8 no significant difference between the methods RK and RV is found. But for the more important correlation estimates, table 4.9 shows that the simple realized covariance is very sensitive to the sampling frequency and generates very deviating estimates. This impression is supported by a t-test and Wilcoxon rank test with a null-hypothesis of mean equal to zero.

		ORCL	IBM	PFE	GOOG	XOM
Realized	tick	0.0130	0.0088	0.0106	0.0110	0.0095
kernel	5 min	0.0129	0.0087	0.0104	0.0110	0.0094
	difference	9.71 (*)	8.31(**)	25.87 (***)	2.72	11.58 (***)
Simple	tick	0.0127	0.0088	0.0130	0.0113	0.0095
realized	5 min	0.0130	0.0088	0.0106	0.0111	0.0095
variance	difference	-25.74 (***)	1.94	-27.20 (***)	19.40 (***)	-3.63

Table 4.8: Averages of realized kernel estimates of daily standard deviation for the five stocks based on tick data and realized kernels based on 5 - minute returns and the average difference in 10^{-5} between these estimates. Equivalently for realized volatility estimates. A t-test was conducted with alternative hypothesis of mean deviation $\neq 0$ with significant p-values in brackets, * for 10%, ** for 5% and *** for 1%.

	ORCL	IBM	PFE	GOOG	XOM
ORCL		-27.78 (***)	-28.04 (***)	-32.61 (***)	-18.57 (***)
IBM	0.24		-22.82 (***)	-24.06 (***)	-17.84 (***)
PFE	2.23 (*) (*)	-0.77		-24.58 (***)	-24.56 (***)
GOOG	2.08 (**)(***)	0.32 (*)	1.03		-21.21 (***)
XOM	-2.51 (*)	-0.28	2.14 (**)	-1.21	

Table 4.9: Mean deviation in 10^{-6} between realized kernel estimates based on tick data and on 5 - minute data in the lower triangular. Equivalently for realized variance in upper triangular. A t-test was conducted with alternative hypothesis of mean deviation $\neq 0$ with significant p-values in brackets, Due to the deviation from normality, results from a Wilcoxon test are given in green (in upper triangula they coincide); * for 10%, ** for 5% and *** for 1%.

4.4 Estimation Methods for the Copula Parameter

For all estimators the Clayton copula is assumed for the dependence structure of the five assets. Copula estimation based on high - frequency data is a new field in statistics and econometrics, and there are some new proposed estimation methods for which implementation will be covered in the next sections.

4.4.1 Method of Moments Type Estimator using Hoeffding's Lemma

The copula dependence parameters θ is estimated by exploiting the information from the covariance of random variables. Here, the integral representation of the covariance of two random variables described in section 3.3 is used.

For the covariance matrix, realized kernel estimates and alternatively the estimated realized covariance was used. As explained in section 3.2, a bivariate distribution together with two marginals is needed. For the joint distribution, the 2 - dimensional Clayton copula replaces the bivariate distribution function and standard normal univariate distributions are taken for the marginals.

$$\sigma_{ij}(\theta) = \int\int [C_\theta\{\Phi(x_i), \Phi(x_j)\} - \Phi(x_i)\Phi(x_j)]dx_i dx_j \tag{4.2}$$

The right part of equation (4.2) does not need real data and only numerical integration must be performed, so that the resulting value depends only on the chosen copula parameter θ.

For all five stocks the estimated 5×5 correlation matrix has 10 different values of $\hat{\sigma}_{ij}$. By applying the minimization as described in section 3.3, an estimate denoted by θ^{MM} is found.

Unfortunately the double integration together with the optimization problem results in a large computational time. The obtained estimates are denoted by θ^{MM} and for simplicity this method is called method of moments estimator.

4.4.2 Estimator Built on Kendall's Tau

Given some estimation of a covariance matrix, values for Kendall's tau can be computed using the transformation rule

$$\tau_{ij,t} = \frac{2}{\pi} \arcsin \rho_{ij,t}$$

that means it is assumed that from the covariance matrix the correlations $\rho_{ij,t}$ between asset i and j are used that are similar to linear correlation.

For many copulae there exists an explicit relationship between Kendall's tau and the copula parameter θ. Genest and Rivest (1993) provided the following form of such a relationship for Archimedean copulae.

$$f_\tau(\theta) = 4 \int_0^1 \phi_\theta^{-1}(\nu)/(\phi_\theta^{-1})'(\nu) d\nu + 1$$

Since the Clayton copula is applied, the mapping function reduces to $f_\tau(\theta) = \frac{\theta}{\theta+2}$ that can be found in table 3.2 and the resulting estimate for a day t is

$$\hat{\theta}_t = \frac{2}{d(d-1)} \sum_{k=1}^{10} f_\tau^{-1}(\hat{\tau}_{k,t}) = \frac{2}{20} \sum_{k=1}^{10} \frac{2\hat{\tau}_{k,t}}{1-\hat{\tau}_{k,t}}$$

where $\hat{\tau}_t$ is obtained by taking all elements from the lower triangular of the estimated correlation matrix for day t and applying the transformation described above.

4.4.3 Realized Dependence Estimator

A Canonical Maximum Likelihood estimation is carried out using the available high frequency data, on the one hand for all available data, that is observations taken mostly every second, and on the other hand 5 - minute returns.

According to the approach in Lidan Großmaß (2013, 2011) a non-parametric ex-post estimate of a daily time - varying copula dependence parameter is found by applying a univariate ARMA(1,1)-GARCH(1,1) process to each of the intraday stock returns. In contrast to the model in Lidan Großmaß (2013), the volatility model regarded here does not contain a part for the conditional mean, that is the term μ_t in the price process$\mathbf{Y_t} = \mu_t + \sigma_t \eta_t$ is set zero.

For each of the five stock returns a GARCH(1,1) model was fit. Vectors of standardized residuals η_{it} with $i = 1, \ldots, 5$ were obtained. For these vectors the empirical cumulative distribution function was estimated and inserted in the formula

$$\hat{\theta} = \max_\theta \sum_{t=1}^{n} \log\, c(\hat{F}_{1,n}(\eta_{1,t}), \ldots, \hat{F}_{5,n}(\eta_{5,t}), \theta)\}$$

where $\hat{F}_{i,n}$ are empirical cdf's and c is the 5 - dimensional density function of the Clayton copula. For this application it was very useful that the Clayton copula was chosen since in contrast to other copula families its density can quite easily be given for 5 dimensions. Maximization delivers the realized dependence estimates as called in Lidan Großmaß (2013, 2011).

4.4.4 Rolling Window Maximum Likelihood Estimator

For this estimator, the inference functions for margins method due to Joe (1997) was applied on daily data taken from Yahoo. Starting from 07.10.2008, a fixed rolling window of 500 days was set from which the marginal distributions of daily log returns were estimated via maximum likelihood. Moreover, a normal distribution with mean equal to zero was fitted to all five stocks by chosing the variance σ^2 that maximized the likelihood. The estimate for the copula dependence parameter was

computed again using the theory of maximum likelihood with

$$\hat{\theta}^{ML} = \arg\max_{\theta} \sum_{t=1}^{T} \left[\sum_{j=1}^{5} \log \phi_{\hat{\sigma}_j}(x_{j,t}) + \log\, c\{\Phi_{\hat{\sigma}_1}(x_{1,t}), \ldots, \Phi_{\hat{\sigma}_5}(x_{5,t}); \theta\} \right]$$

where ϕ denotes the normal density function, Φ the normal distribution and $x_{i,t}$ is the daily log return for the i'th asset and day t, $i = 1, \ldots, 5$.

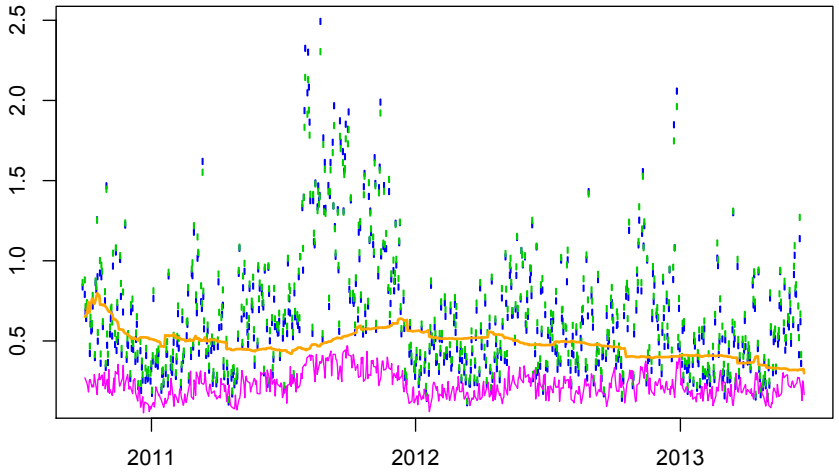

Figure 4.4.1: Based on tick data, estimated realized dependence parameter (magenta line) together with θ^{MM}(blue points), $\theta^{Ad\,hoc}$(green points) and θ^{ML} (orange line), where θ^{MM} and $\theta^{Ad\,hoc}$ are estimated using RK.

Having computed all estimates, some graphical analysis was done to compare these four estimators. Furthermore some descriptive statistics are given in table 4.10.

The resulting estimates for the realized dependence have more in similar with the estimates of θ^{MM} and $\theta^{Ad\,hoc}$ when the RV estimates are based for the covariance. This becomes clear in figure 4.4.2, where in contrast to figure 4.4.1, all estimates lie more or less together. By replacing RK with RV for estimating θ^{MM} and $\theta^{Ad\,hoc}$ these two estimators yield smaller values below 1. With other words, θ^{MM} and $\theta^{Ad\,hoc}$ are reduced to the level of the realized dependence estimates where the ML copula parameter estimates θ^{ML} are placed. This change can also be found in table 4.7, where the correlations in the RV matrix are almost half the size of those from RK. Going to the lower sampling frequency of 5 - minutes should increase the level of the values from the θ^{RD} estimates. This is verified in figure 4.4.3 in which the realized kernel estimates were used for θ^{MM} and $\theta^{Ad\,hoc}$. Also, in table 4.10, it can be seen, that realized dependence estimates using 5 - minute data have larger values and higher variation than with tick data. The copula parameter estimates are very similar for 5 - minute returns if instead of RK the simple realized variance and covariance is used.

As another comparison of estimates based on tick and 5 - minute data, the figure 4.4.4 shows the tail dependence coefficient that is given for the Clayton copula with parameter θ by $\lambda_L = 2^{-1/\theta}$ ignoring the fact that for some estimators the marginal distributions are assumed to be normal. Since

the Clayton copula generates lower tail dependence, it was quantified in this figure 4.4.4.

		Min	Mean	Median	Max	Std. deviation
tick	θ^{MM}	0.11	0.64	0.55	2.49	0.39
	$\theta^{Ad\,hoc}$	0.12	0.68	0.60	2.30	0.37
	θ^{RD}	0.05	0.23	0.22	0.47	0.08
5-minute returns	θ^{MM}	0.11	0.68	0.55	3.30	0.44
	$\theta^{Ad\,hoc}$	0.12	0.75	0.64	2.94	0.43
	θ^{RD}	0.04	**0.53**	**0.49**	**1.58**	**0.25**
daily	θ^{ML}	0.30	0.48	0.49	0.80	0.08

Table 4.10: Descriptive statistics of the four copula dependence parameter estimates, for tick data and 5 - minute returns. The estimators θ^{MM} and $\theta^{Ad\,hoc}$ are obtained by using RK.

		Min	Mean	Median	Max	Std. deviation
tick	θ^{MM}	0.09	0.66	0.55	3.17	0.39
	$\theta^{Ad\,hoc}$	0.10	0.71	0.62	2.82	0.37
5-minute returns	θ^{MM}	0.05	0.28	0.26	0.71	0.12
	$\theta^{Ad\,hoc}$	0.06	0.31	0.28	0.78	0.13

Table 4.11: Descriptive statistics of the copula dependence parameter estimates, for tick data and 5 - minute returns. The estimators θ^{MM} and $\theta^{Ad\,hoc}$ are obtained by using RV.

Having estimated the copula parameter θ for each day with different estimation methods, it can be used together with the realized kernel estimates for comparing the resulting copula models by examination of the Value-at-Risk.

4.5 Value-at-Risk

Value-at-Risk (VaR) are frequently used to quantify risk exposures. This estimates represent a critical value of a portfolio's potential profit and loss distribution for one day, as explained in Kupiec (1995). The Value-at-Risk estimation requires the availability of extreme observation, since it deals with extreme quantiles. Such observations are rare so that the VaR estimates ca be incorrect. Therefore it is important to assess and quantify the accuracy of VaR estimates. In the following the construction of the VaR estimate is explained and afterwards the results are tested on accuracy including a comparison relating to the different copula parameter estimators.

Regarding a linear portfolio of d positions, its value is defined as the weighted sum of the individual prices $S_t = (S_{1,t}, \ldots, S_{d,t})^\top$ at time $t \in [1, T]$,

$$V_t = \sum_{i=1}^{d} w_i S_{i,t}$$

with the weights $w = (w_1, \ldots, w_d)^\top$. The profit and loss (P&L) function is formed as

$$L_{t+1} = V_{t+1} - V_t = \sum_{j=1}^{d} w_j S_{j,t} \left\{\exp(X_{j,t+1}) - 1\right\}$$

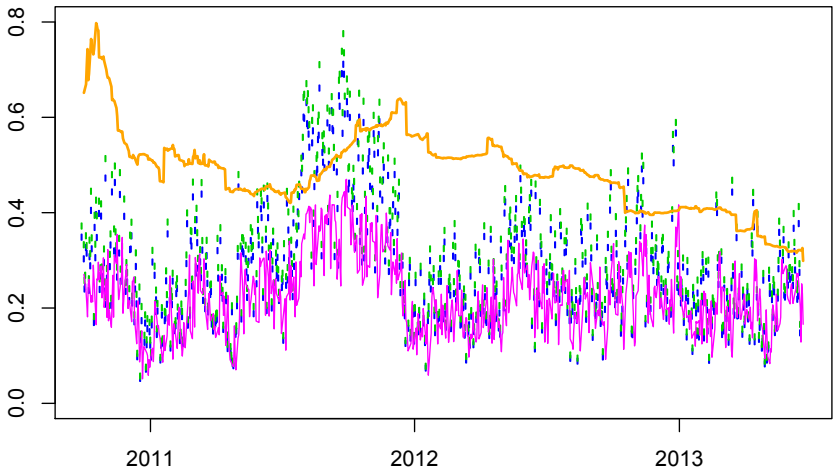

Figure 4.4.2: Estimates of Lidan's θ^{RD}(magenta line), the daily θ^{ML}(orange line), θ^{MM}(blue points) and $\theta^{Ad\,hoc}$(green points), where the last two estimators are based on the estimated RV.

where $X_{t+1} = \log S_{t+1} - \log S_t$ are the log-returns. The Value-at-Risk is the α- quantile of the P&L distribution, $VaR(\alpha) = F_L^{-1}(\alpha)$.

For the P&L function, daily data from Yahoo was taken. From the 684 observations, log-returns for 683 days were available for each asset.

For evaluating the copula models based on the four different parameter estimators θ^{MM},$\theta^{Ad\,hoc}$, θ^{RD} and $\theta^{ML\,daily}$, the loss distribution of a portfolio of $d = 5$ positions, ORCL, IBM, PFE, GOOG and XOM was simulated from a Clayton copula distribution with sample size 1000 using the realized kernel estimates for σ for the normal marginals and the estimated copula dependence parameter $\hat{\theta}$.

$$C_{\hat{\theta}}\left\{\Phi_{\hat{\sigma}_1}(X_1), \ldots, \Phi_{\hat{\sigma}_5}(X_5)\right\}$$

In addition, the traditional realized volatility and realized covariance estimator was used, both combined in the notation RV.

By this, for each day there were 1000 simulated values for the loss distribution and different α-quantiles were used for VaR estimation. For the α- level, 1%, 5% and 10% were taken.

In order to asses the performance of the four copula estimation methods, some backtesting was done. A backtest serves for evaluating a risk model, in this case the the model on which the VaR estimation is based. Using the daily adjusted closing prices from Yahoo, the true realizations $\{l_t\}$ of the P&L function are computed and compared to the estimated Value-at-Risk values. To measure the quality of the estimated $\widehat{VaR}_t(\alpha)$, a backtest proposed by Kupiec (1995) was applied, which is an unconditional coverage test. It focuses on the exceedance ratio $\hat{\alpha}$, which gives the percentage of cases, where the true loss l_t was smaller than the calculated $\widehat{VaR}_t(\alpha)$.

$$\hat{\alpha} = \frac{1}{T}\sum_{t=1}^{T} I\{l_t < \widehat{VaR}_t(\alpha)\}$$

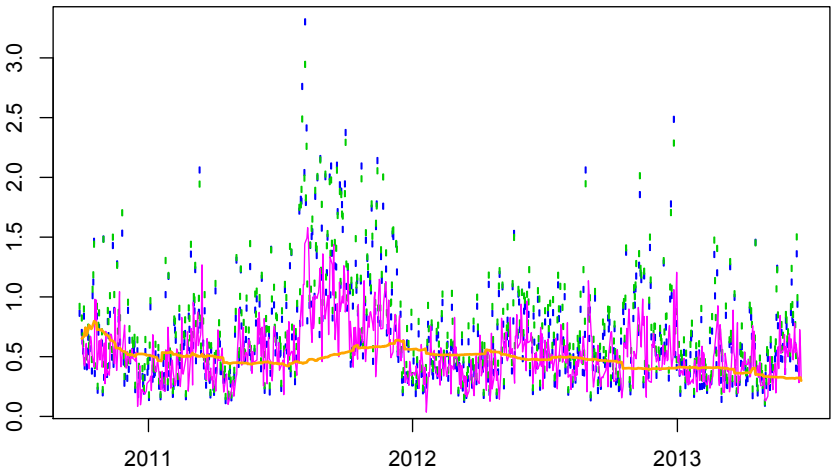

Figure 4.4.3: Copula parameter estimates for θ^{RD}(magenta line), the daily θ^{ML}(orange line), θ^{MM}(blue points) and $\theta^{Ad\,hoc}$(green points). 5 - minute returns were used for estimating θ^{RD}, $\theta^{MM}$and $\theta^{Ad\,hoc}$ and RK was used for $\theta^{MM}$and $\theta^{Ad\,hoc}$.

where T is the number of available values of $\widehat{VaR}_t(\alpha)$ and $N = T \cdot \hat{\alpha}$ is the absolute number of observed exceedances. Kupiec's test statistic is defined as

$$K(p) = 2 \cdot \log \left\{ \left(\frac{1-\hat{p}}{1-p} \right)^{T-N} \left(\frac{\hat{p}}{p} \right)^{N} \right\}$$

Since the realized dependence estimator and the maximum likelihood estimator are constructed on a different model than θ^{MM} and $\theta^{Ad\,hoc}$, it was interesting to check the Value-at-Risk by simulating from the respective correct model. However, the resulting VaR values were worse for the daily maximum likelihood model when the variance was not taken from the RK estimates but from maximum likelihood estimates based on daily data.

Furthermore, Value-at-Risk estimation with copula has higher efficiency and flexibility than other methods based on a normality assumption (see Franke et al. (2011) chapter 17).

From the tables 4.12 and 4.13 can be seen, that the 1% quantile is better estimated for θ^{ML} and θ^{RD} when using realized kernel estimates for the simulation from the Clayton copula. However, it must be mentioned that by drawing new Monte Carlo samples these result could change slightly.

4.6 HAR Forecasting

In this section, Value-at-Risks are calculated by using forecasts of the dependence parameters.

In order to capture the dynamics of the copula dependence parameter, the Heterogeneous autoregressive model (HAR) discussed in Andersen et al. (2007) and Corsi (2009) was implemented, since it is relatively easy to estimate. In this model the conditional volatility is modeled with a special type of linear model that can be estimated by ordinary least squares. In Corsi (2009) this is described as an additive cascade model of volatility components. These components are realized over different time

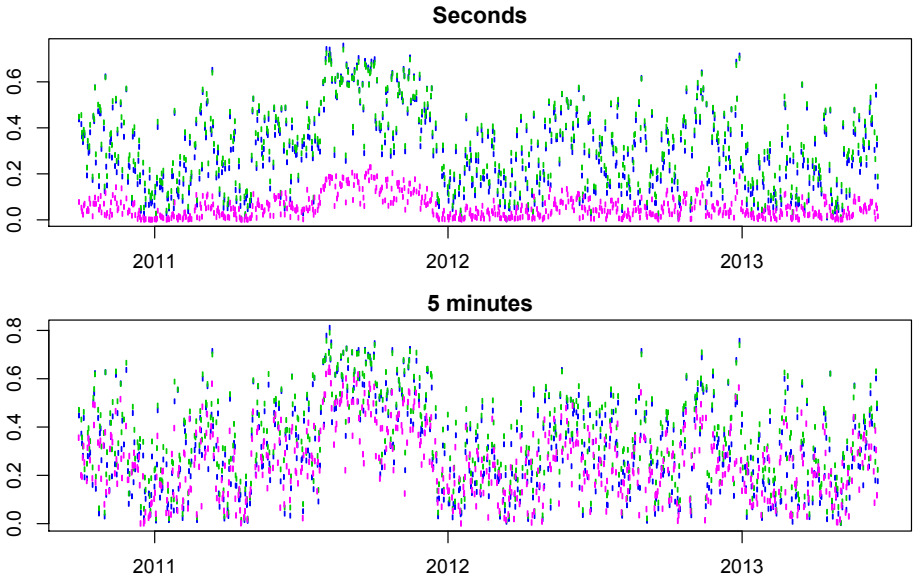

Figure 4.4.4: Lower tail dependence coefficient for the Clayton copula constructed by θ^{MM} (blue dots), $\theta^{Ad\,hoc}$(green dots) and θ^{RD}(magenta dots). In the upper panel estimation was based on all available data (variable sampling frequency around 1 second) and in the lower panel only 5 - minute returns were used for estimation.

horizons to imitate different market behavior.

As a result, the variance is forecasted one day ahead. Additionally, the copula parameter is also forecasted using the HAR model, as explained in Fengler and Okhrin (2012). Thus, 6 models were estimated

$$\begin{aligned}
\log(\hat{h}_{1,t+1|t}) &= \beta_0^1 + \beta_D^1 \log\hat{\sigma}_t^D + \beta_W^1 \log\hat{\sigma}_t^W + \beta_M^1 \log\hat{\sigma}_t^M \\
\dots &= \dots \\
\log(\hat{h}_{5,t+1|t}) &= \beta_0^5 + \beta_D^5 \log\hat{\sigma}_t^D + \beta_W^5 \log\hat{\sigma}_t^W + \beta_M^5 \log\hat{\sigma}_t^M \\
\hat{\hat{\theta}}_{t+1|t} &= \alpha_0 + \alpha_D\hat{\theta}_t^D + \alpha_W\hat{\theta}_t^W + \alpha_M\hat{\theta}_t^M
\end{aligned}$$

where $\hat{\sigma}_i$ is the RK estimate of the variance of stock i, $\hat{\theta}_t$ is an estimate of the copula parameter for day t and the superscripts D, W and M indicate daily, weekly and monthly averages, that is $\theta_t^D = \theta_t$, $\theta_t^W = \frac{1}{5}\sum_{i=0}^{4}\theta_{t-i}$ and $\theta_t^M = \frac{1}{21}\sum_{i=0}^{20}\theta_{t-i}$. This procedure is illustrated in figure 4.6.1. For estimating the HAR model, the first 250 days beginning from 01.10.2010 were used, such that time series of forecasts contained 434 days.

From the previous sections it could be seen, that the realized kernel is a good estimator for the variation in the data. Thus a time series of the variances of the five stocks was taken from the realized kernel estimates. It is assumed, that the variance $\hat{\sigma}_{t,1}, \dots, \hat{\sigma}_{t,5}$ is time dependent and standard time series models can be applied.

The forecasted values for the variance and the copula parameter were used to construct a Clayton copula, from which Monte Carlo samples were drawn. Again, Value-at-Risk estimates were obtained, that are given in table 4.14.

For IBM the forecasted time series is presented together with the estimates in figure 4.6.2.

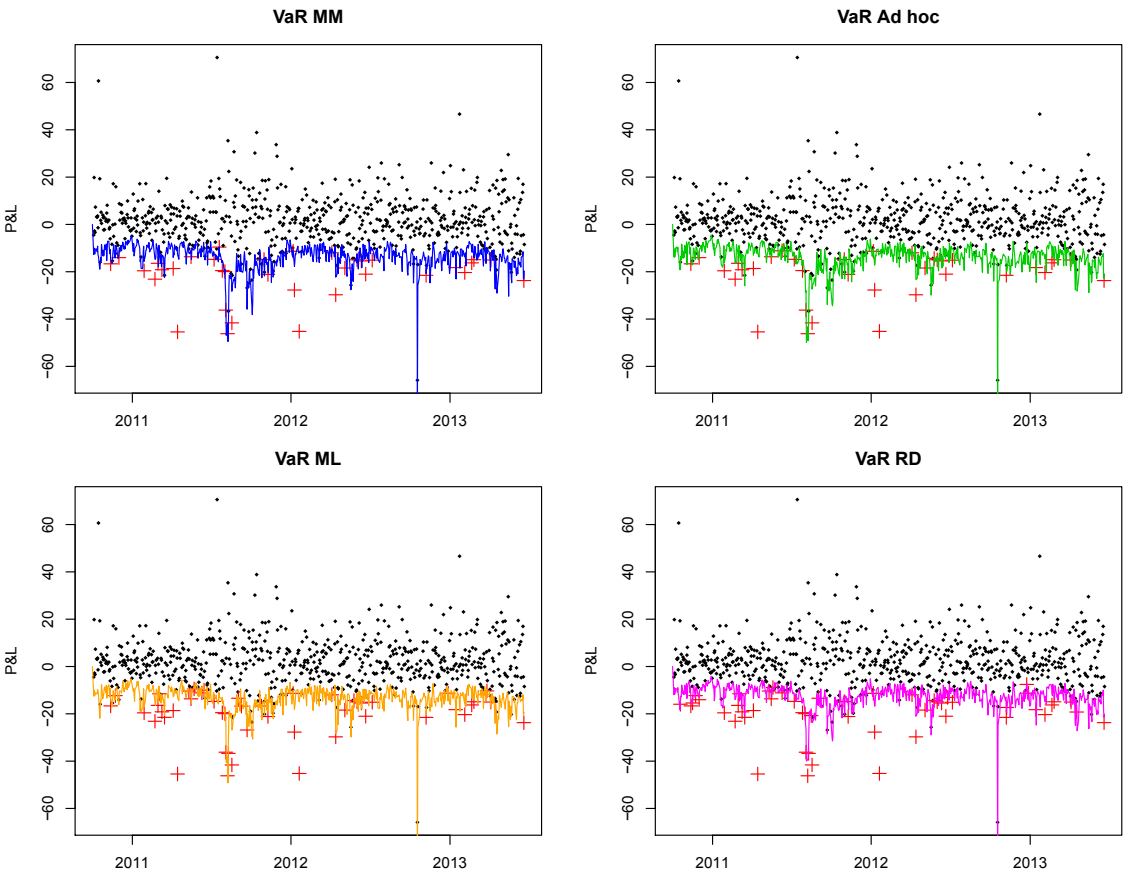

Figure 4.5.1: Value-at-Risk estimates, $\widehat{VaR}_t(\alpha)$ (solid line), P&L (dots) and exceedances (red crosses), for $\alpha = 0.05$.

		1 second returns		
	Estimator	VaR(0.01)	VaR(0.05)	VaR(0.10)
RK	θ^{MM}	**1.03 (0.948)**	**4.98 (0.979)**	11.86 (0.114)
	θ^{Adhoc}	1.17 (0.661)	**5.12 (0.882)**	11.13 (0.334)
RV	θ^{MM}	1.17 (0.661)	6.30 (0.135)	13.18 (0.008)
	θ^{Adhoc}	1.32 (0.426)	6.30 (0.135)	13.32 (0.006)
RK for simulation	θ^{RD}	1.32 (0.426)	7.91 (0.001)	14.64 (0.000)
	θ^{ML}	**1.02 (0.948)**	6.15 (0.183)	12.74 (0.022)
RV for simulation	θ^{RD}	1.47 (0.254)	**6.44 (0.97)**	13.47 (0.004)
	θ^{ML}	1.32 (0.426)	5.71 (0.405)	13.32 (0.006)

Table 4.12: Value-at-Risk exceedance rates in % for θ^{MM}, θ^{Adhoc}, θ^{RD} and θ^{ML}. VaR was computed from simulated returns using a Clayton copula with according dependence parameter and realized kernel estimates. The Kupiec test is presented in brackets.

		5-minute returns		
	Estimator	VaR(0.01)	VaR(0.05)	VaR(0.10)
RK	θ^{MM}	**1.02 (0.948)**	5.71 (0.405)	12.74 (0.022)
	θ^{Adhoc}	0.88 (0.744)	**5.12 (0.882)**	11.86 (0.114)
RV	θ^{MM}	0.88 (0.744)	4.69 (0.703)	12.59 (0.029)
	θ^{Adhoc}	**1.02 (0.948)**	**5.12 (0.882)**	12.59 (0.029)
RK for simulation	θ^{RD}	**1.03 (0.948)**	6.30 (0.135)	13.03 (0.011)
	θ^{ML}	1.17 (0.661)	7.17 (0.014)	13.03 (0.011)
RV for simulation	θ^{RD}	1.17 (0.661)	**5.12 (0.882)**	12.45 (0.039)
	θ^{ML}	1.46 (0.254)	6.59 (0.069)	12.59 (0.029)

Table 4.13: Value-at-Risk exceedance rates for θ^{MM}, θ^{Adhoc}, θ^{RD} and θ^{ML}. VaR was computed from simulated returns using a Clayton copula with according dependence parameter and realized kernel estimates. The Kupiec test is presented in brackets.

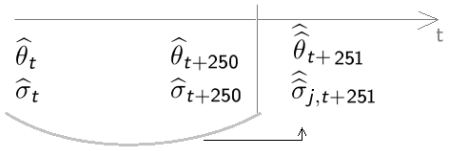

Figure 4.6.1: One-day ahead forecasting of variance and copula parameter. A double hat represents the forecasts and a single hat the estimates.

Estimator	VaR(0.01)	VaR(0.05)	VaR(0.10)
θ^{MM}	2.54 (0.007)	6.70 (0.125)	11.78 (0.235)
θ^{Adhoc}	2.08 (0.049)	6.70 (0.125)	**11.08 (0.486)**
θ^{RD}	3.00 (0.001)	8.31 (0.004)	13.39 (0.025)
θ^{ML}	2.77 (0.002)	7.62 (0.020)	12.24 (0.136)

Table 4.14: Value-at-Risk exceedance rates in % for forecasted values of θ^{MM}, θ^{Adhoc}, θ^{RD} and θ^{ML}. VaR was computed from simulated returns using a Clayton copula with forecasted dependence parameter and forecasted variances.

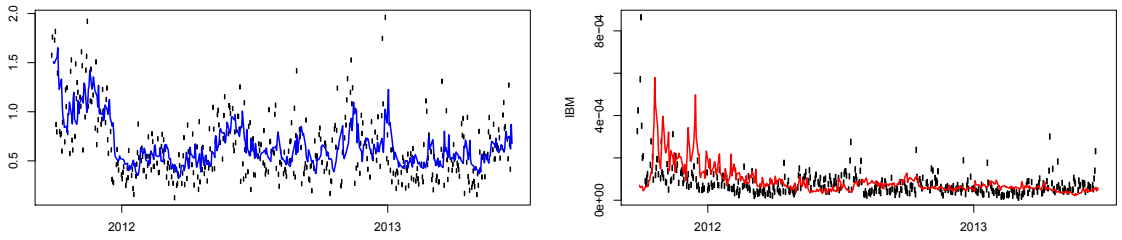

Figure 4.6.2: Forecasts for the daily variance (lower panel) of IBM and the copula parameter θ^{MM}(upper panel)

Chapter 5

Conclusions

The goal of this thesis is to apply and examine up-to-date approaches for estimating copula models based on intraday financial data. For estimation of dependence between risk factors like stock returns involving copula models, in the last years an increased number of publications has appeared. In contrast, the new approach of estimating the daily ex-post dependence between assets by making use of high-frequency data is quite new and some of the few proposed estimation methods, that from Fengler and Okhrin (2012) and Lidan Großmaß (2013) were applied. Four copula parameter estimators are built on different settings, parametric and semi-parametric.

In the ad hoc estimation, the non-parametric rank statistic Kendall's tau offered a rather simple and fast way for estimation the dependence parameter. A key for this is the available variability information. The high-frequency data type has different properties than those of daily data and this requires special estimation methods for the covariance between assets. The popular realized volatility and realized covariance estimators appeared in this work to be inferior to the realized kernel estimator, which accounts for the autocorrelation in the returns and is less sensitive to market microstructure effects.

As the last estimator, maximum likelihood estimator uses only daily price data and due to the rolling window the parameter estimates change over time. Dependence estimates from this estimator differ greatly from the results of the other estimators because the dependence in the portfolio of the five stocks looks more or less constant and shows no reaction to events on the world market like financial crisis.

The main object in modeling stock returns is to accurately describe the true return distribution and to be able to get precise estimates of extreme quantiles which describe possible profits or losses in times of a financial boom or crisis. Therefore, all estimated copula dependence parameter were used in the combination with Value-at-Risk calculation.

Confirming the present literature on copula estimation methods, the method of moments estimator using Hoeffding's lemma together with the ad hoc estimator delivered the best results. It must be stressed, that these estimators performed best when the realized kernel estimator was chosen for estimating covariance. When the standard realized variance, as the sum of squared intraday returns, was used to estimate variability in the data, the Value-at-Risk performance fell of in quality. Furthermore, using only 5 - minute returns reduced the intraday data amount and led to slightly imprecise quantile estimation

High frequency data is an enrichment for researchers in the sense that risk modeling is facilitated. As pointed out in Barndorff-Nielsen et al. (2011) and also in Bollerslev et al. (2008), high frequency

data is essential for evaluating and forecasting risk factors.

Sometimes assets behave mostly independent in normal periods, but move together in crisis. This was observed for data of stock returns in this work.

Using the out-of-sample forecast of the method of moments estimator, the estimated Value-at-Risk was found to perform better than when other copula dependence estimators were used.

The realized dependence estimator of Lidan Großmaß (2013) performed well for 5 - minute returns and led to much worse results for the returns with a higher sampling frequency of about seconds. This can be due to microstructure effects or / and the irregular spaced observations. All in all, the realized dependence estimator was outdone by the method of moments estimator and the ad hoc estimator for all sampling frequencies.

Bibliography

Andersen, T. and T. Bollerslev: 1998, 'Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts'. *International Economic Review* **39**(4), 885–905. Symposium on Forecasting and Empirical Methods in Macroeconomics and Finance.

Andersen, T., T. Bollerslev, and F. Diebold: 2007, 'Roughing it up: including jump components in the measurement, modeling and forecasting of return volatility'. *Review of Economics and Statistics* **79**, 701–720.

Andersen, T., T. Bollerslev, F. Diebold, and P. Labys: 2001, 'The Distribution of Realized Exchange Rate Volatility'. *Journal of the American Statistical Association* **96**, 42–55.

Barndorff-Nielsen, O., P. R. Hansen, A. Lunde, and N. Shephard: 2008a, 'Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise.'. *Econometrica* **76**, 1481–1536.

Barndorff-Nielsen, O., P. R. Hansen, A. Lunde, and N. Shephard: 2008b, 'Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading.'. *Working paper*. Institut for Okonomi, Aarhus Universitet.

Barndorff-Nielsen, O., P. R. Hansen, A. Lunde, and N. Shephard: 2009, 'Realized kernels in practice: trades and quotes.'. *Econometrics Journal, Royal Economic Society* **12**, C1–C32.

Barndorff-Nielsen, O., P. R. Hansen, A. Lunde, and N. Shephard: 2011, 'Multivariate realized kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading.'. *Journal of Econometrics* **162**, 149–169.

Barndorff-Nielsen, O. E. and N. Shephard: 2004, 'A Feasible Central Limit Theory for Realised Volatility Under Leverage'. Economics Papers 2004-W03, Economics Group, Nuffield College, University of Oxford.

Bollerslev, T., G. Tauchen, and H. Zhou: 2008, 'Expected Stock Returns and Variance Risk Premia'. AFA 2008 New Orleans Meetings Paper; Review of Financial Studies, Forthcoming; Duke Department of Economics Research Paper No. 5; CREATES Research Paper No. 2008-48.

Brandimarte, W. and D. Bases: 2011, 'United States loses prized AAA credit rating from S&P'. Reuters.

Chen, X., Y. Fan, and V. Tsyrennikov: 2006, 'Efficient Estimation of Semiparametric Multivariate Copula Models'. *Journal of the American Statistical Association* **101**(475), 1228–1240. Theory and Methods.

Chen, Y., W. K. Härdle, and U. Pigorsch: 2010, 'Localized Realized Volatility Modeling'. *Journal of the American Statistical Association* **105:492**, 1376–1393.

Christensen, K., S. Kinnebrock, and M. Podolskij: 2010, 'Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data'. *Journal of Econometrics* **159**, 116–133.

Corsi, F.: 2009, 'A Simple Approximate Long-Memory Model of Realized Volatility'. *Journal of Financial Econometrics* **7**(2), 174–196.

Duan, J.-C., W. Härdle, and J. E. Gentle: 2011, *Handbook of Computational Finance*. Springer.

Fengler, M. and O. Okhrin: 2012, 'Realized copula'. working paper.

Franke, J., W. K. Härdle, and C. M. Hafner: 2011, *Statistics of Financial Markets, An Introduction*. Springer, 3rd edition.

Genest, C. and A.-C. Favre: 2007, 'Everything You always wanted to know about copula modeling but were afraid to ask'. *Journal of Hydrologic Engineering* **12**, 347–368.

Genest, C., K. Ghoudi, and L.-P. Rivest: 1995, 'A semiparametric estimation procedure of dependence parameters in multivariate families of distributions'. *Biometrika* **82**, 543–552.

Genest, C. and L.-P. Rivest: 1993, 'Statistical Inference Procedures for Bivariate Archimedean Copulas'. *Journal of American Statistical Association* **88**, 1034–1043.

Hafner, C. M. and H. Manner: 2008, 'Dynamic stochastic copula models: Estimation, inference and applications'. (043).

Härdle, W. K. and O. Okhrin: 2009, 'De copulis non est disputandum, Copulae: An Overview'. working paper.

Hautsch, N., L. M. Kyj, and R. C. Oomen: 2009, 'A blocking and regularization approach to high dimensional realized covariance estimation'. Discussion paper.

Jin, X. and J. M. Maheu: 2013, 'Modeling realized covariances and returns'. *Journal of Financial Econometrics* **11**(2), 335–369.

Joe, H.: 1997, *Multivariate Models and Dependence Concepts*. Chapman & Hall.

Kendall, M. and J. D. Gibbons: 1990, *Rank correlation methods*. Oxford University Press, USA, 5 edition.

Kupiec, P.: 1995, 'Techniques for Verifying the Accuracy of Risk Management Models'. *Journal of Derivatives* **3**, 73–84.

Lidan Großmaß, R.: 2011, 'Using Intraday Data for Estimation of Multivariate Dependence'. working paper.

Lidan Großmaß, R.: 2013, 'Three Essays on Using High Frequency Data in Estimating Financial Risks'. PhD dissertation, Universität Konstanz.

McNeil, A. J., R. Frey, and P. Embrechts: 2005, *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton Series in Finance. Princeton University Press.

Nelsen, R. B.: 2007, *An Introduction to Copulas*. Springer, second edition.

Oh, D.-H. and A. J. Patton: 2011, 'Modelling dependence in high dimensions with factor copulas'. Working paper, Duke University.

Oh, D.-H. and A. J. Patton: 2013, 'Simulated method of moments estimation for copula-based multivariate models'. *Journal of the American Statistical Association*.

R Core Team: 2013, 'R: A Language and Environment for Statistical Computing'. R Foundation for Statistical Computing, Vienna, Austria.

Sklar, A.: 1959, *Fonctions de répartition á n dimensions et leurs marges*. Université Paris 8, 5 edition.

Taylor, S. J. and X. Xu: 1997, 'The incremental volatility information in one million foreign exchange quotations'. *Journal of Empirical Finance* **4**, 317–340.

Trivedi, P. K. and D. M. Zimmer: 2005, 'Copula Modeling: An Introduction for Practitioners'. *Foundations and Trends in Econometrics* **1**(1), 1–111.

Erklärung

Hiermit erkläre ich, Tatjana Tissen-Diabaté, dass ich die vorliegende Arbeit allein und nur unter Verwendung der aufgeführten Quellen und Hilfsmittel angefertigt habe.

Die Prüfungsordnung ist mir bekannt. Ich habe in meinem Studienfach bisher keine Masterarbeit eingereicht bzw. diese nicht endgültig nicht bestanden.

Berlin, den Unterschrift: _____